

**Ursachen der Bewertung des Reiseerlebens – Eine Mehrebenen-Konfirmatorische Faktorenanalyse der
Satisfaction with Travel Scale**

Masterarbeit

im Studiengang M.Sc. Psychologie mit dem Schwerpunkt kognitive Neurowissenschaften

von Stefan Bohmann

Matrikelnummer: 228989

am Institut für Sozial- und Persönlichkeitspsychologie an der
Otto-von-Guericke Universität Magdeburg^a

und Institut für Verkehrssystemtechnik
des Deutschen Zentrums für Luft- und Raumfahrt e.V.^b

Erstgutachterin: Dr. Esther J. Bosch^b

Zweitgutachter: Prof. Dr. Florian G. Kaiser^a

Abgabedatum: 19.11.2024

Zusammenfassung

Diese psychometrische Studie untersucht die Satisfaction with Travel Scale (STS) zur Bewertung tatsächlich erlebter, multimodaler Reisen im öffentlichen Nahverkehr. Im Rahmen einer Experience-Sampling Studie absolvierten die Probanden sechs bis zehn vorgegebene Routen. Die unbalancierten Längsschnittdaten der 42 analysierten Probanden wurden erfolgreich mithilfe einer Mehrebenen-konfirmatorischen Faktorenanalyse in *R lavaan* modelliert. Nach notwendigen Modellmodifikationen konnte das Mehrebenenmessmodell die im realen Kontext erhobenen Messwiederholungsdaten erfolgreich reproduzieren ($\chi^2(df) = 30.81(27)$ $p = .279$, $_{Robust}RMSEA = .021$). Explorative Korrelationsanalysen der modellierten Faktoren mit den berichteten Ereignissen der Reise implizieren, zusammen mit den vorgenommenen Modellmodifikationen, dass zukünftige Versionen der STS zusätzliche Items zur kognitiven Bewertung der Reise entwickeln und weniger des Affekts verwenden sollten. Die Ursachen der Bewertung des Reiseerlebens lassen sich als Funktion der erlebten Reise und der bewertenden Person beschreiben.

Keywords: Satisfaction with Travel Scale, Satisfaction, Multilevel Confirmatory Factor Analysis, Travel Experience, Public Transport

Hinweis zur geschlechtergerechten Sprache

Aufgrund einer erleichterten Lesbarkeit wird in dieser Arbeit das generische Maskulinum verwendet, wobei sämtliche Geschlechter gleichermaßen mitgemeint sind, sofern nicht explizit konkreter angegeben.

Inhaltsverzeichnis

Einleitung	1
Stand der Forschung	1
Satisfaction with Travel Scale (STS)	6
Konstruktdefinition und Ziel der Studie.....	8
Methoden	10
Stichprobenbeschreibung	10
Durchführung.....	12
Messungen.....	13
Mehrebenen-Konfirmatorische Faktorenanalyse (MCFA) in einer Nussschale.....	15
Auswertung.....	19
Ergebnisse	20
Deskriptiv	20
Die Reise zum Mehrebenenmessmodell	23
Mehrebenenmessmodell (MMM)	27
Interpretation der Faktorenstruktur der STS.....	29
Explorative Analysen.....	30
Reisefriedenheit und Verkehrsmittelpräferenz	34
Diskussion	36
Literaturverzeichnis	45
Anhang A.....	55
Anhang B	56
Eigenständigkeitserklärung.....	60

Einleitung

Eine Reduktion des motorisierten Individualverkehrs kann maßgeblich dazu beitragen, die energiebedingten globalen CO₂-Emissionen zu reduzieren und die Lebensqualität in Städten durch Faktoren wie Luftverschmutzung, Flächennutzung und Lärmbelästigung zu verbessern (Lee et al., 2023; Monteiro, Sousa, Coutinho-Rodrigues & Natividade-Jesus, 2024). Der öffentliche Personen-Nahverkehr (ÖPNV) stellt eine umweltfreundlichere Alternative dar. Dies wirft die Frage auf, welche Faktoren die Verkehrsmittelwahl bestimmen.

Stand der Forschung

De Vos, Singleton und Gärling (2022) beschreiben den Travel Mode Choice Cycle als einen zyklischen Prozess, bei dem die Wahl des Verkehrsmittels die Bewertung der Reise (Reisezufriedenheit) beeinflusst, die wiederum eine verkehrsmittelbezogene Einstellung, in ihrem Sinne eine Meinung über das jeweilige Verkehrsmittel, prägt. Eine positive Einstellung gegenüber einem Verkehrsmittel erhöhe die Wahrscheinlichkeit, dieses Verkehrsmittel erneut zu benutzen. Letztere Beziehung wird im Sinne der Theory of Planned Behavior (TPB; Ajzen, 1991) über die Intention zur Verkehrsmittelnutzung mediiert und durch Variablen wie der subjektiven Norm, also dem wahrgenommenen sozialen Druck ein Verhalten (nicht) zu zeigen, moderiert. Lim, Thompson, Pearson, Caldwell Odgers und Beck (2024) erweitern das Modell um das subjektive Erleben während einer Fahrt (Reiseerleben), das affektive und kognitive Bewertungen (Appraisals; Scherer, 2001; 2019) verschiedener Stimuli während einer Fahrt, wie zum Beispiel Überfüllung oder Verspätungen, umfasst (Lim, Thompson & Beck, 2023).

Der Mode Choice Cycle bietet eine konzeptuelle Übersicht der potenziell relevanten Variablen im Kontext des Reiseerlebens. Die Zusammenhänge zwischen diesen Variablen sowie ihre Konzeptualisierung benötigen jedoch weiterer Forschung. Der einzige im Zuge des Mode Choice Cycles angeführte empirische Beleg für die Wirkung der Reisezufriedenheit auf eine verkehrsmittelbezogene Einstellung liefert eine explorative Querschnittsanalyse von De Vos, Schwanen, van Acker und Witlox (2019). Die in dieser Studie

berichteten geringen direkten Effekte von .08 bei Fahrradfahrern und .1 bei ÖPNV-Nutzern lassen sich möglicherweise auf einen *common method bias* zurückführen, da beide Konstrukte mittels Likert-Skalen erfasst wurden (Podsakoff, MacKenzie, Lee & Podsakoff, 2003). Zudem basieren diese Schätzungen auf ungewichteten Mittelwerten anstelle messfehlerbereinigter Faktorscores.

Schwache empirische Belege für den Zusammenhang theoretischer Konstrukte wie Reisezufriedenheit und einer verkehrsmittelbezogenen Einstellung deuten nicht zwangsläufig darauf hin, dass ein solcher Zusammenhang nicht existiert. Unzulängliche Erhebungsinstrumente können aufgrund von Messfehlern und unvollständiger Erfassung eines Konstrukts die tatsächlichen Zusammenhänge zwischen Variablen verschleiern oder abschwächen (Wilson, 2023). Die Qualität eines Messinstruments ist untrennbar an die Güte der zugrunde liegenden Theorie gebunden. Wenn die Theorie die Realität unzureichend abbildet, kann das darauf basierende Instrument keine präzisen und validen Messungen liefern.

Die im Kontext des Mode Choice Cycles genutzten Einstellungsmaße basieren größtenteils auf der TPB und erheben Einstellung vorwiegend anhand meinungsbasierter Likert-Skalen. Gemäß des Tripartite Model of Attitude (Rosenberg & Hovland, 1960) sind affektive und kognitive Reaktionen sowie Verhaltensweisen bezüglich eines Einstellungsobjekts, also sowohl Meinungen als auch Gewohnheiten, Manifestationen einer Einstellung. Im Gegensatz zur TPB bietet das sparsamere Campbell-Paradigma, das ebenfalls aus dem Tripartite Model herleitbar ist, ein verhaltensbasiertes Einstellungskonstrukt (Kaiser, Byrka & Hartig, 2010; Kaiser, 2021). Dieses Rasch-Modell (Rasch, 1960) erklärt Zielverhaltensweisen durch ihre spezifischen Verhaltenskosten, wie dem Aufwand persönlicher Ressourcen (zum Beispiel Geld, Zeit, Soziale Anerkennung), die durch das persönliche Engagement einer Person bezüglich des Ziels, respektive ihrer Einstellung, kompensiert werden. Die General Ecological Behavior Scale (GEB) misst Umwelteinstellung auf theoretischer Grundlage des Campbell-Paradigmas anhand selbstberichteten Verhaltens (Kaiser, 1998; Kaiser, 2020). Das zugrundeliegende Messmodell ist Rasch-skaliert und bietet

somit höchste Präzision (z.B. Yu, 2020) sowie eine unkomplizierte Auswertung. Die Entscheidung zwischen umweltfreundlicheren Verkehrsmitteln, wie Fahrrad oder ÖPNV und dem umweltschädlicheren Auto kann durch die Ausprägung der Umwelteinstellung und den situativen Verhaltenskosten erklärt werden (Kaiser, Kibbe & Hentschke, 2021). Nachhaltiges Mobilitätsverhalten entspricht sogar einer Facette des GEB.

Zwar sollte eine Umwelteinstellung nicht die Entscheidung zwischen ähnlich umweltfreundlichen Verkehrsalternativen wie Bus und Straßenbahn vorhersagen können, doch diese Unterscheidung ist zur breitflächigen Reduzierung des motorisierten Individualverkehrs weniger bedeutsam. Angesichts der sparsamen, nützlichen und psychometrisch hochwertigen Operationalisierung einer verhaltensbasierten Einstellung durch das Campbell Paradigma erscheinen meinungsbasierte Einstellungstheorien wie die TPB und darauf basierende Instrumente in der verkehrspsychologischen Forschung überrepräsentiert.

Um die Wahl zu umweltfreundlicheren Verkehrsmitteln auf breiter Ebene zu fördern, sollten Interventionen und Anpassungen der gebauten Umwelt als permanente Lösungen geplant werden, beispielsweise fahrradfreundliche Infrastruktur und vergünstigte Fahrttickets (wie das Deutschland-Ticket). Kurzfristige Interventionen, wie monetäre Anreize zur Förderung spezifischer Verhaltensweisen, wirken nur für die Dauer der Interventionen, aber nicht über die geförderte Verhaltensweise und Laufzeit hinaus (Kaiser, Henn & Marschke, 2020; Steinhorst & Klöckner, 2018).

Die weitere Möglichkeit, die sich aus dem Campbell-Paradigma ergibt, ist eine Einstellungsänderung, also eine Stärkung der intrinsischen Motivation, sich umweltbewusst zu Verhalten. Eine stärkere Motivation manifestiert sich in allen Verhaltensweisen, indem höhere Verhaltenskosten aller umweltfreundlichen Verhaltensweisen in Kauf genommen wären (Henn, Otto & Kaiser, 2020). Die Wahrscheinlichkeit jeder einzelnen Verhaltensweise stiege nur geringfügig, doch kumuliert über alle möglichen umweltfreundlichen Verhaltensweisen wäre die Wirkung beachtlich.

Vor der Fragestellung, wie die intrinsische Motivation gefördert werden kann, erscheint die postulierte Wirkung der Reisezufriedenheit auf eine verkehrsmittelbezogene Einstellung sehr vielversprechend. Überzeugende empirische Belege für diesen Mechanismus bleiben jedoch aus. Wie bereits erwähnt, sollte dieser Zusammenhang erst verworfen werden, wenn die Konstrukte überzeugend gemessen werden und die Erhebungsmethode angemessen ist, um die Fragestellung zu beantworten. Dass die tatsächliche Nutzung öffentlicher Verkehrsmittel mit dem Reiseerleben zusammenhängt, erscheint zumindest plausibel.

Unter der Annahme, dass angenehme Erlebnisse verstärkend und unangenehme Erlebnisse dagegen bestrafend wirken, legen die behavioristischen Prinzipien des menschlichen Verhaltens, insbesondere der Operanten Konditionierung (Domjan, 2013; Skinner, 1938), nahe, dass positive Erlebnisse im ÖPNV die Wahrscheinlichkeit erhöhen, dass dieses Verkehrsmittel erneut gewählt wird, unangenehme Erlebnisse diese jedoch verringern, wie es das Law of Effect von Thorndike (1933) bereits impliziert.

Kahneman und Kollegen untersuchten in einer Reihe von Studien Heuristiken zur Bewertung erlebter Episoden, vor allem den Zusammenhang zwischen Bewertungen des Erlebens im jeweiligen Moment, der globalen Bewertung der erlebten Episode und den aus Erlebnissen resultierenden Präferenzen (z.B. Kahneman, 2000; Kahneman & Frederick, 2005). Beispielsweise bewerteten Patienten während Koloskopien alle 60 Sekunden auf einer Likert-Skala ihre aktuelle Schmerzintensität (Redelmeier & Kahneman, 1996). Nach Abschluss bewerteten sie die gesamte Prozedur. Interessanterweise zeigte sich kein Zusammenhang zwischen der Dauer der Koloskopie und ihrer globalen Bewertung ($r = .03$). Vielmehr war eine Linearkombination aus der höchsten Schmerzbewertung und der letzten Bewertung der beste Prädiktor der Gesamtbewertung. Die daraus abgeleitete *Peak-End Heuristik* fand in einer Folgestudie Anwendung. Dabei wurde einer klinischen Experimentalgruppe, die sich einer Koloskopie unterzogen, ohne medizinische Notwendigkeit die Prozedur um eine Minute verlängert. In dieser zusätzlichen Minute

wurde das Instrument nicht bewegt, damit das Ende der Prozedur weniger schmerzhaft erlebt wurde. Die Experimentalgruppe zeigte eine signifikant höhere globale Bewertung ihrer Koloskopie im Vergleich zur Kontrollgruppe, die keine verlängerte Prozedur erfuhr. Besonders beeindruckend ist, dass nach Berücksichtigung von Störvariablen wie der individuellen Krankheitsgeschichte die Patienten der Experimentalgruppe mit einer geringfügig erhöhten Wahrscheinlichkeit zu Folgeuntersuchungen einige Jahre später erschienen (Odds Ratio = 1.41) (Redelmeier, Katz & Kahneman, 2003). Obwohl ein solcher *Extension Neglect* zur Bewertung erlebter Episoden, in denen die Dauer ein salientes Merkmal ist, wie bei Fahrten mit dem ÖPNV, nicht zwingend zu erwarten ist (Ariely & Loewenstein, 2000), könnten solche unintuitiven Manipulationen des Erlebens ein noch unausgeschöpftes Potenzial zur Verhaltensänderung bieten.

Aufbauend auf den Arbeiten zu Heuristiken der globalen Bewertung erlebter Episoden untersuchten Abenoza, Cats & Susilo (2019) explorativ, wie sich die retrospektive Bewertung multimodaler ÖPNV-Fahrten aus den Bewertungen ihrer einzelnen Abschnitte zusammensetzt. Die Ergebnisse der 130 getesteten Modelle deuten eher gegen heuristische Aggregationsregeln wie der Peak-End Regel und stattdessen für ein zeitlich gewichtetes Aggregat der Bewertungen aller Abschnitte einer Fahrt. Allerdings stellen die Bewertungen der einzelnen Abschnitte bereits aggregierte Bewertungen dar, die teilweise retrospektiv erhoben wurden. Außerdem verwendeten sie, wie Redelmeier und Kollegen (1996, 2003) nur ein einzelnes Likert-Skalen-Item zur Erfassung der Zufriedenheit. Ein-Item Messungen sind inhärent unreliabel (z.B. Weidman, Steckler und Tracey, 2017). Die klassischen Methoden der Psychometrie bieten keine Möglichkeit zu überprüfen, was mit diesen Items wirklich gemessen wird, da keine lösbaren Modellgleichungen aufgestellt werden können. Um verworfene Hypothesen über Reisezufriedenheit und dessen Zusammenhänge mit weiteren Variablen aufzustellen, bedarf es zunächst eines überprüfbaren Messinstruments.

Satisfaction with Travel Scale (STS)

Neben Ein-Item-Zufriedenheitsmessungen ist die Satisfaction with Travel Scale (STS) das am häufigsten verwendete Instrument zur Erfassung der Reisezufriedenheit verschiedener Verkehrsmittel (Ettema et al. 2011; Friman, Fujii, Ettema, Gärling & Olsson, 2013). Reisezufriedenheit meint gemäß der jüngsten Erklärung durch De Vos und Kollegen (2022) die erlebten Emotionen während einer Reise sowie die kognitive Evaluation der gesamten Reise oder einzelner Teilabschnitte. Reisezufriedenheit sei zeitlich daran gebunden, wie lange eine Person über die Reise nachdenkt oder die erlebte Episode als Gedächtnisinhalt abrufbar ist. In Abgrenzung zu einer Meinung über bestimmte Verkehrsmittel oder dem Reisen allgemein ist Reisezufriedenheit somit ein situatives Konstrukt.

Die hierarchische Faktorenstruktur der STS impliziert, dass sich die Zufriedenheit mit einer Fahrt als Faktor 2. Ordnung in einer kognitiven und affektiven Bewertung äußert. Die affektive Bewertung lässt sich gemäß des Affekt-Circumplex Modells (Russell, 1980) in eine positive Deaktivierung (bzw. negative Aktivierung) und eine positive Aktivierung (bzw. negative Deaktivierung) differenzieren, die jeweils durch Faktoren 1. Ordnung repräsentiert sind. Diese manifestieren sich in jeweils drei Items auf 7-stufigen Skalen, abgeleitet aus der Swedish Affect Core Scale (Västfjäll, Friman, Gärling & Kleiner, 2002). Die Extrema der Skalen sind jeweils mit Adjektiven beschriftet, die Gefühle mit niedriger oder hoher Valenz sowie niedrigem oder hohem Arousal beschreiben sollen, zum Beispiel *gehetzt (1)* bis *entspannt (7)* (positive Deaktivierung). Die kognitive Bewertung als weiterer Faktor 1. Ordnung manifestiert sich ebenfalls in drei Items, die eine geringe bis hohe Zufriedenheit beschreiben sollen, zum Beispiel *Die Reise hat schlecht funktioniert (1)* bis *Die Reise hat Reibungslos funktioniert (7)*. Innerhalb dieser Konzeptualisierung muss bedacht werden, dass die Beantwortung der affektiven Items selbst ein kognitiver Prozess ist, der verschiedene Informationsverarbeitungsschritte wie das Verstehen der Frage, dem Abruf von Gedächtnisinhalten und einer Urteilsbildung beinhaltet (Borg & Staufenbiehl, 2007 S. 38). Die Verwendung von Adjektiven als Itemanker birgt das Risiko, dass verschiedene Personen diese

unterschiedlich verstehen oder ihr Selbstbericht durch die verwendeten Begriffe verzerrt ist (Barrett, 2006). Da die kognitive Bewertung des erlebten Affekts zum Zeitpunkt der Erhebung geschieht, ist diese Bewertung höchstens als Aggregat der Dynamik des Affekts während der Fahrt zu verstehen.

Singleton (2019) liefert eine Übersicht psychometrischer Analysen der STS und verwendet fünf affektive und vier kognitive Items in seiner Variante des Fragebogens. Die Ergebnisse seiner Konfirmatorischen Faktorenanalysen (CFA), um den Zusammenhang zwischen dem Messinstrument und dem zu messenden Konstrukt zu überprüfen und zu quantifizieren, unterstützen eine hierarchische Faktorenstruktur ($\chi^2(df) = 99.05(24)$, $CFI = .967$, $SRMR = .034$, $RMSEA = .081$), im Einklang mit den meisten psychometrischen Studien der STS (Olsson et al, 2012; Friman et al., 2013; Glasgow, Geller, Le & Hankey, 2018). Zwei Studien reproduzierten ihre erhobenen Daten besser mit einem zweifaktoriellen Messmodell (De Vos, Schwanen, Van Acker & Witlox, 2015; Smith, 2017). In einigen dieser psychometrischen Studien konnte eine akzeptable Modellgüte allerdings erst nach Modellmodifikationen, wie das Zulassen korrelierender Messfehler zwischen Items, erreicht werden, die teilweise nicht näher erläutert wurden (z.B. Friman et al. 2013). In keiner dieser Studien wurde die STS zum Abschluss einer Fahrt erhoben, sondern es erfolgten retrospektive Bewertungen der zuletzt absolvierten, einer üblichen oder hypothetischen Fahrt.

Ein möglicher Grund für die starke Überrepräsentation retrospektiver Querschnitts-Onlinebefragungen liegt vermutlich im geringeren Aufwand im Vergleich zu Labor- und Feldstudien, wodurch die Akquise hunderter Probanden erst möglich ist. Solche Querschnittsbefragungen bieten bereits wertvolle Einblicke in das Konstrukt der Reisezufriedenheit, insbesondere in dessen Faktorenstruktur. Allerdings verwässern solche einmaligen, retrospektiven Erhebungen die Grenze zwischen der Reisezufriedenheit als situatives Konstrukt und einer allgemeinen Zufriedenheit mit Reisen oder einer Meinung über Reisen. Wie dargelegt, unterliegt vermutlich bereits eine Bewertung zum Abschluss einer erlebten Episode Heuristiken. Je länger der Zeitraum zwischen der Reise und dem

Bewertungszeitpunkt ist, desto stärker wird der Abruf der Episode durch Erinnerungsfehler beeinflusst (Shapira & Pansky, 2019). Es stellt sich die Frage, wie viel der Reisezufriedenheit wirklich auf die erlebte Reise oder auf die bewertende Person zurückzuführen ist.

Daten aus Längsschnitterhebungen ermöglichen im Vergleich zu Querschnittsdaten eine detaillierte Analyse der stabilen und fluktuierenden Eigenschaften der Reisezufriedenheit über verschiedene Situationen hinweg (Hoffman, 2015). Durch wiederholte Messungen werden mehr Informationen pro Proband erfasst, was kleineren Stichproben entgegenwirkt. Allerdings verletzen solche Daten aufgrund der Abhängigkeiten einzelner Messzeitpunkte über die gemessenen Personen die Annahmen der klassischen CFA. Um dennoch faktorenanalytisch zu untersuchen, wie sich das zu messende Konstrukt Reisezufriedenheit in dem Messinstrument STS manifestiert, ist eine erweiterte Methode der CFA erforderlich, die Mehrebenen-Konfirmatorische Faktorenanalyse (MCFA; Muthén, 1994). Diese berücksichtigt nicht nur die Abhängigkeit der Messzeitpunkte, sondern ermöglicht erst eine Analyse der situations- und personenspezifischen Eigenschaften des Konstrukts Reisezufriedenheit.

Konstruktdefinition und Ziel der Studie

Eine Reise umfasst die Bewegung von einem Startort zu einem Zielort zu Fuß oder durch Nutzung eines oder mehrerer Verkehrsmittel sowie Phasen des Wartens. Reiseerleben meint die subjektiven Wahrnehmungen während einer Reise und umfasst sowohl kognitive und affektive Reaktionen auf äußere Reize (Ereignisse während der Reise) als auch innere, unbewusste Prozesse und Dynamiken des Affekts und der Kognition einer reisenden Person. Reisebezogene Aspekte außerhalb der zeitlichen und räumlichen Episode einer Reise, wie der Buchungsprozess, können ebenfalls Teil des Reiseerlebens sein.

Im Kontext dieser Studie bezeichnet Reisezufriedenheit eine abschließende Bewertung des Reiseerlebens. Diese ergibt sich aus der Bewertung zum Abschluss einer Reise, im Gegensatz zu retrospektiven oder Bewertungen während einer Reise. Die STS ist das Instrument zur Erfassung des Konstrukts Reisezufriedenheit. Diese erfasst die Bewertung des Reiseerlebens mit Indikatoren zur

kognitiven Bewertung der Reise und des erlebten Affekts zum Ende der Reise. Darin verbirgt sich die Annahme, dass der zum Erhebungszeitpunkt erlebte Affekt durch die unmittelbar zuvor erlebte Reise bestimmt wird.

Ein psychometrisch fundiertes Instrument zur Erfassung der Bewertung des Reiseerlebens erlaubt es, Hypothesen über Zusammenhänge mit weiteren Konstrukten oder Aspekten der gebauten Umwelt zu prüfen. Auch wenn dadurch kein beeindruckender Mechanismus zur Verhaltens- oder Einstellungsänderung offenbart wird, ist die Bewertung des Reiseerlebens ein wichtiges Konstrukt, da eine vermehrte ÖPNV Nutzung zur Reduktion des motorisierten Individualverkehrs ein öffentliches Interesse darstellt (Bundesministerium für Digitales und Verkehr, 2023). Dementsprechend sollte es für Stakeholder, Verkehrsbetriebe und Politik von zentralem Interesse sein, das Nutzungserleben im ÖPNV möglichst angenehm zu gestalten. Ein Instrument zur validen Erfassung der Bewertung des Reiseerlebens ist notwendig, um belastbare Daten über die tatsächlich unangenehmen Aspekte des Reiseerlebens zu generieren und darauf aufbauend wirksame Verbesserungsmaßnahmen abzuleiten.

Diese Studie untersucht die STS als Bewertung tatsächlich erlebter, multimodaler ÖPNV-Fahrten, erhoben im Rahmen einer Experience-Sampling Studie, in der Versuchsteilnehmer eine festgelegte Strecke wiederholt fahren und bewerteten. Es ist kein Ziel der Studie die Faktorenstruktur STS zu bestätigen oder zu verwerfen. Vielmehr soll sie zum Verständnis der Konstrukte Reiseerleben und dessen Bewertung, der Reisezufriedenheit, beitragen. Durch die Anwendung der MCFA wird Reisezufriedenheit sowohl auf der Personen- als auch auf der Situationsebene analysiert, um ein umfassendes Verständnis des Konstrukts zu erlangen.

Methoden

Die Experience-Sampling (ES) Studie zur Erfassung des Reiseerlebens wurde am Institut für Verkehrssystemtechnik des Deutschen Zentrums für Luft- und Raumfahrt e.V. (DLR) von Januar bis Dezember 2023 im Rahmen des Projekts DATAMOST durchgeführt (Bosch, Luther & Ihme, 2024) und von der DLR- Ethikkommission geprüft und genehmigt (Antrags-Nummer 7/22).

Stichprobenbeschreibung

Über das DLR Probandentool und einen Aushang am Schwarzen Brett der TU Braunschweig wurden 45 deutschsprachige Probanden (mindestens C1-Niveau) rekrutiert. Drei Probanden wurden ausgeschlossen, weil sie jeweils höchstens zwei Fahrten wie instruiert durchführten. Tabelle 1 beschreibt die 42 analysierten Probanden. Die Studienteilnahme wurde mit 60-100€ vergütet, je 10€ pro abgeschlossener Studienfahrt. Die Fahrtkosten wurden erstattet, falls keine Dauerkarte vorhanden war. Fünf der Probanden absolvierten die vorgesehene Studienstrecke als Teil ihres regulären Pendelwegs.

Tabelle 1*Soziodemographische Eigenschaften der Stichprobe*

Eigenschaft	<i>n</i>	%
Geschlecht		
Weiblich	21	50
Männlich	21	50
Bildungsabschluss		
Kein Schulabschluss	1	2
Realschule (mittlere Reife)	4	10
(Allgemeine) Hochschulreife	23	55
Studienabschluss	14	33
Wohnsituation		
Ländlich	1	2
Stadt bis 50.000 Einwohner	4	10
Stadt bis 100.000 EW	4	10
Stadt bis 500.000 EW	32	76
Größere Stadt	1	2
Führerscheinbesitz	35	83
Arbeitsweg		
Kürzer als 1 km	5	12
1-5 km	18	43
5-10 km	7	17
10-20 km	6	14
Länger als 20 km	6	14
Fortbewegungsmittel zur Arbeit^a		
Laufen	20	48
Radfahren	23	55
Bus	27	64
Zug	13	31
Auto (Fahrer)	8	19
Straßenbahn	9	21

Bemerkungen. *N* = 42. Probanden waren durchschnittlich 29 Jahre alt (*SD* = 13.8, Median = 24.5, Spanne 19 bis 79).

^aMehrfachauswahl möglich

Durchführung

Die Probanden wurden umfassend über die Ziele der Studie zur Erfassung des Reiseerlebens im öffentlichen Personennahverkehr, die Datenschutzbestimmungen, die Freiwilligkeit ihrer Teilnahme sowie den Ablauf der Studie informiert. Sie erhielten am Institut einen Elektrokardiogramm-Gurt und ein Smartphone zur Erfassung der Herzfrequenz, GPS-Daten und Experience Sampling Daten (movisens GmbH, 2023) während der Fahrten sowie der Fragebögen zu Beginn und am Ende der Studie.

Innerhalb von zwei Wochen absolvierten die Probanden mindestens sechs bis maximal zehn vorgegebene Routen mit verschiedenen Verkehrsmitteln zwischen Braunschweig (ca. 250.000 Einwohner) und Wolfenbüttel (ca. 50.000 Einwohner). Jede Studienfahrt umfasste wenigstens die folgenden Abschnitte (Abbildung 1):

Abbildung 1

Bus- und Zugrouten der Studienfahrt zwischen Braunschweig und Wolfenbüttel



1. Reise von zu Hause zu einer Bushaltestelle in Braunschweig mit einem frei gewählten Verkehrsmittel. Aus Datenschutzgründen sollten die Probanden die Datenaufnahme erst beginnen, nachdem sie ihre Wohnstraße verlassen haben.
2. Weiterfahrt mit dem Bus zur Haltestelle Rathaus
3. Weiterfahrt mit der Straßenbahn zum Braunschweiger Hauptbahnhof
4. Fahrt nach Wolfenbüttel, entweder mit dem Bus (50% der Fahrten) oder mit dem Zug (restliche 50% der Fahrten)
5. Fußweg zum Stadtzentrum von Wolfenbüttel

Die Rückfahrt erfolgte in umgekehrter Reihenfolge und zählte als separate Fahrt. Die Probanden konnten ihre Reise in Braunschweig oder Wolfenbüttel beginnen. Die Aufzeichnungen der Fahrten zum und vom Institut zählten nicht als Studienfahrten, konnten aber freiwillig aufgezeichnet werden.

Messungen

Zu Beginn füllten die Probanden Fragebögen zur Demographie und den BFI-10 (10-Item Big-Five Inventory; Rammstedt, Kemper, Klein, Beierlein & Kovaleva, 2013) aus. Dieser erfasst die Big-Five Persönlichkeitsfaktoren (Quelle) mit jeweils zwei Items auf einer Likert-Skala von 1 (trifft überhaupt nicht zu) bis 5 (trifft voll und ganz zu). (1) Neurotizismus – „Ich werde leicht nervös und unsicher“, „Ich bin entspannt, lasse mich durch Stress nicht aus der Ruhe bringen (-)“; (2) Extraversion – „Ich gehe aus mir heraus, bin gesellig“, „Ich bin eher zurückhaltend, reserviert“ (-); (3) Offenheit – „Ich habe eine aktive Vorstellungskraft, bin phantasievoll“, „Ich habe nur wenig künstlerisches Interesse“ (-); (4) Verträglichkeit – „Ich erledige Aufgaben gründlich“, „Ich bin bequem, neige zu Faulheit“ (-); (5) Gewissenhaftigkeit – „Ich schenke anderen leicht Vertrauen, glaube an das Gute im Menschen“, „Ich neige dazu, andere zu kritisieren“ (-).

Während der Fahrten wurden die Probanden alle 3,5 Minuten per Alarm des Smartphones aufgefordert, einen ES-Fragebogen auszufüllen. Dieser umfasst Fragen zur aktuellen Zufriedenheit mit der Reise, zur Stimmung und zum empfundenen Stress auf Likert-Skalen (1-10) sowie zur offenen Formulierung der aktuell empfundenen Emotion. Zusätzlich wurden sie gefragt, ob seit der letzten

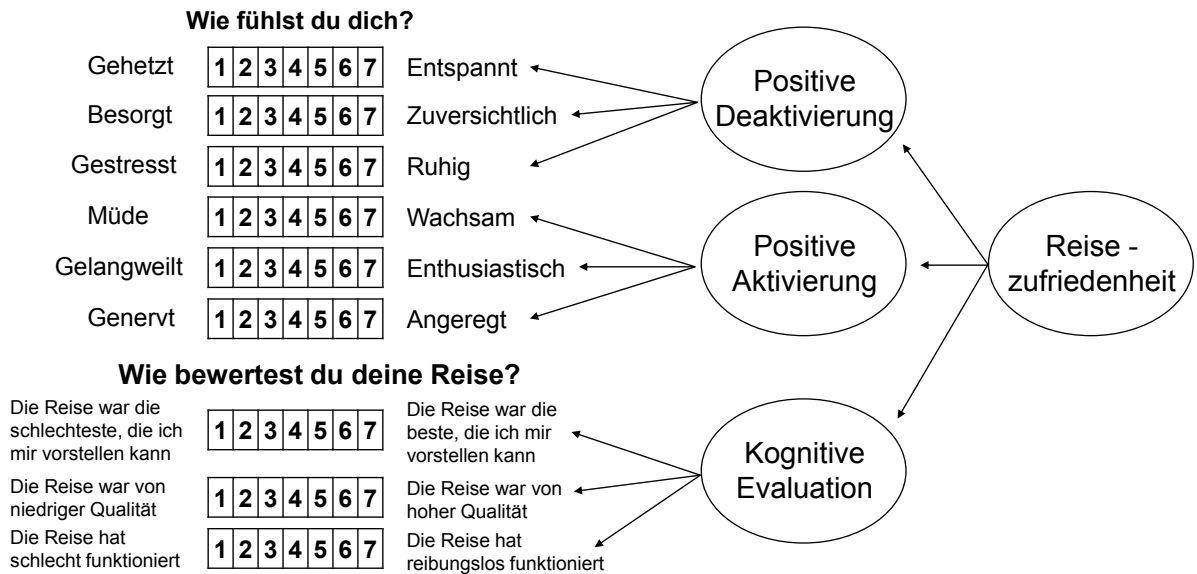
Abfrage bestimmte Ereignisse eingetreten seien, mit einer offenen und folgenden vorgegebenen

Antwortoptionen:

1. „Verspätung / Anschluss verpasst / Beeilung nötig“
2. „Störende Menschen in der Umgebung / Überfülltes Verkehrsmittel“
3. „Fahrverhalten des Fahrers / der Fahrerin“
4. „Infrastruktur (z.B. keine Überdachung / keine oder nicht genügend Informationen /nicht barrierefrei / unebene Straße)“
5. „Positive soziale Situation / Interaktion“
6. „Mediale Unterhaltung (Musik, Podcast, Buch, etc.)“
7. „Ziel oder Anschluss erreicht / Verkehrsmittel pünktlich“
8. „Unwohlsein (Langeweile, Ungeduld, schwierige Gedanken, Hunger, Müdigkeit, Frieren, Nässe, etc.)“
9. „Verkehrsmittel ist komfortabel (gemütlich / angenehme Temperatur / nicht zu voll / Sitzplatz bekommen etc.)“
10. „Schöne Umgebung“ und „Private / Berufliche Erledigungen“

Die Probanden konnten die Alarme ablehnen und waren instruiert, eigenständig den ES-Fragebogen aufzurufen, sobald etwas berichtenswertes geschehen ist oder das Verkehrsmittel gewechselt wurde.

Zum Abschluss jeder Fahrt wurde die STS mit neun Items (in Deutsch übersetzt aus Ettema et al., 2011; Olsson et al., 2013) erhoben (Abbildung 2). Nach Abschluss aller Fahrten berichteten die Probanden von ihren erlebten Reisen in einem Abschlussfragebogen. Zudem gaben sie an, ob sie für künftige Reisen zwischen Braunschweig und Wolfenbüttel den Bus oder den Zug präferierten und begründeten ihre Wahl im offenen Antwortformat.

Abbildung 2*Satisfaction with Travel Scale (STS)***Mehrebenen-Konfirmatorische Faktorenanalyse (MCFA) in einer Nusschale**

Die CFA ist ein statistisches Verfahren zur Überprüfung von Messmodellen (Jöreskog, 1969; 1971; Campbell & Fiske, 1959). Ein Messmodell beschreibt die theoretisch angenommene Beziehung zwischen beobachtbaren Variablen (Indikatoren) und nicht direkt beobachtbaren, latenten Variablen (Faktoren) als multiple Regression. Demnach äußert sich ein Faktor in verschiedenen Indikatoren und erklärt deren Kovarianz. Respektive kovariieren die Indikatoren, weil sie durch einen gemeinsamen Faktor verursacht werden (Hoyle, 2012; Moosbrugger & Kelava, 2020). Die übrige Varianz der Indikatoren, die nicht durch den Faktor erklärt wird, wird entsprechend der klassischen Testtheorie (Gulliksen, 1950; Steyer & Eid, 2001) als Messfehler betrachtet. Das Ziel eines Messmodells ist eine möglichst sparsame und korrekte Repräsentation der Varianzen und Kovarianzen von Indikatoren auf Populationsebene. Zur Überprüfung der Modellgüte wird die Differenz der modellimplizierten Kovarianzmatrix Σ , die sich aus den jeweils geschätzten Modellparametern ergibt und der empirischen Kovarianzmatrix S der Indikatoren gebildet. Je

geringer die Differenz der Matrizen ist, desto besser passt das Modell zu den beobachteten Daten. Mit Ockhams Rasiermesser in der metaphorischen Hand ist jenes Messmodell zu bevorzugen, das mit der geringsten Anzahl an Modellparametern die beobachteten Daten hinreichend reproduzieren kann (z.B. Sober, 1981). Mithilfe der CFA können die theoretischen Annahmen hinter Messungen überprüft und um ihren Messfehler korrigiert werden.

Eine strikte Annahme der Regressionsanalyse und der CFA sind unkorrelierte Residuen der Beobachtungen (z.B. Aitken, 1936). Abhängigkeiten der Beobachtungen können beispielsweise bei Datenerhebungen mit Schülern (Within) an verschiedenen Schulen (Between) (Dunn et al., 2015) oder bei wiederholt beobachteten Situationen (Within) verschiedener Personen (Between) (Duncan et al., 1997; Kyriazos, 2019) auftreten. In solchen Fällen sind die Residuen der Messungen innerhalb der Schule oder Person (Within) oft miteinander korreliert (Hoffman & Walters, 2022). Solche über ihre Cluster korrelierten Residuen bzw. Abhängigkeiten untergraben die Zufallsstichprobe. Das Ignorieren dieser Abhängigkeiten oder ein Aggregieren über die Cluster kann irreführende Zusammenhänge auf aggregierter Ebene offenbaren, wie beispielsweise durch das Simpson-Paradoxon ausgiebig erläutert (Blyth, 1972; Pearl, 2022).

Die Anwendung der CFA auf geclusterte Daten führt zu falsch geschätzten Modellparametern und nicht sinnvoll interpretierbaren Teststatistiken (Julian, 2001; Muthen & Satorra, 1995). Muthen (1989, 1994) zeigte mit der MCFA eine elegante Lösung, um die Abhängigkeiten zu berücksichtigen und die Faktorenstrukturen der Within-Ebene und Between-Ebene gleichzeitig in einem Mehrebenen-Messmodell (MMM) zu überprüfen. Die modellimplizierte Kovarianzmatrix Σ_T wird zerlegt in die Within-Kovarianzmatrix Σ_W und die Between-Kovarianzmatrix Σ_B . Σ_W und Σ_B sind orthogonal und additiv zu Σ_T . Dies gilt ebenso für die zugrundeliegenden Within- und Between Faktoren, weshalb eine äquivalente Faktorenstruktur über die verschiedenen Ebenen des Messmodells nicht unbedingt zu erwarten ist und

die Faktoren jeweils ihre eigene Interpretation im Kontext ihrer Messung erfordern. Die empirische Kovarianzmatrix S_T wird ebenfalls in ihre ebenenspezifischen Varianzanteile S_{PW} und S_B zerlegt:

$$S_T = (N - 1)^{-1} \sum_{g=1}^G \sum_{i=1}^{N_g} (y_{gi} - \bar{y})(y_{gi} - \bar{y})' \quad (1)$$

$$S_{PW} = (N - G)^{-1} \sum_{g=1}^G \sum_{i=1}^{N_g} (y_{gi} - \bar{y}_g)(y_{gi} - \bar{y}_g)' \quad (2)$$

$$S_B = (G - 1)^{-1} \sum_{g=1}^G N_g (\bar{y}_g - \bar{y})(\bar{y}_g - \bar{y})' \quad (3)$$

G = Clustergröße (Anzahl der Personen bei Messwiederholungen)

N = Anzahl der Beobachtungen

S_T wird in der herkömmlichen CFA verwendet und ist im Mehrebenenfall ein konsistenter Schätzer von $\Sigma_T = \Sigma_B + \Sigma_W$. Die gepoolte Within-Kovarianzmatrix S_{PW} ist ein konsistenter und verlässlicher Schätzer von Σ_W (Winer, Brown & Michels, 1991). S_{PW} und S_B sind nicht orthogonal.

S_B ist ein konsistenter und verlässlicher Schätzer von $\Sigma_W + c \Sigma_B$, wobei der Skalierungsfaktor c im balancierten Fall die Clustergröße der Between-Ebene darstellt. Bei ungleicher Clustergröße und einer großen Anzahl an Clustern ist c ungefähr die mittlere Clustergröße. Die Maximum-Likelihood Schätzung von Σ_B ist $c^{-1} (S_B - S_{PW})$ (Muthén, 1989). Deshalb wird der Between-Faktor im Messmodell mit \sqrt{c} skaliert und die Between-Kovarianzmatrix entsprechend zur Schätzung angepasst.

Auf dieser kovarianzanalytischen Grundlage können mithilfe der MCFA die Faktorenstrukturen verschiedener Ebenen von Messungen überprüft werden und Abhängigkeiten in den Daten berücksichtigt werden. Konzeptuell lässt sich veranschaulichen, dass die Faktoren der jeweiligen Ebenen gemeinsam als Ursache der beobachteten Messungen wirken, allerdings statistisch unabhängig voneinander sind. Sie lassen sich durch die präsentierten Kovarianzmatrizen S_{PW} und $c S_B$ schätzen.

Die Notwendigkeit für eine Mehrebenen-Analyse wird an den Intraklassenkorrelationen (*ICCs*) ersichtlich, die den Anteil der Between-Cluster Varianz an der gesamten Varianz beschreibt (Koch, 1983)

Es gibt keine klaren Empfehlungen für Grenzwerte der *ICCs*, ab wann eine Mehrebenenmodellierung zwingend erforderlich ist. Bereits sehr kleine *ICCs* über Null zeigen an, dass Varianz durch Unterschiede zwischen den Clustern erklärt werden kann. Je höher der Anteil der Between-Cluster Varianz an der Gesamtvarianz ist, desto geringer wäre der tatsächliche Informationsgehalt der Within-Messungen. Und desto stärker würde eine unangepasste Stichprobengröße die effektive Stichprobengröße überschätzen und Parameterschätzungen aufblähen. Zur Veranschaulichung dieses Problems kann man sich ein Messwiederholungsdesign vorstellen, in dem 100 Personen jeweils fünf Mal gemessen werden. Eine CFA ohne Berücksichtigung des Clusterings würde mit 500 unabhängigen Beobachtungen rechnen. Angenommen, die Messungen unterscheiden sich zwar zwischen den Personen, aber jede Person gibt in allen fünf Messungen die gleichen Antworten. Da diese sich nicht unterscheiden, liefern sie die gleiche Menge unabhängiger Informationen wie bei einmaliger Messung jeder Person. Die effektive Stichprobengröße wäre also 100 und nicht 500. Dieser Extremfall soll das Problem verdeutlichen, was bereits bei geringen *ICCs* zu Annahmenverletzungen und Fehlinterpretationen führt. Daher ist es wichtig, hierarchisch strukturierte Daten mit Methoden der Mehrebenenmodellierung wie der MCFA korrekt zu modellieren, um valide und interpretierbare Ergebnisse zu erzielen (Snijders & Bosker, 2012).

In der Praxis erweist sich die Schätzung von MMM oft als anspruchsvoll, da sie anfällig für Schätzprobleme wie negative Varianzen (Hox, 2002; Muthén & Asparouhov, 2011) und ausbleibende Konvergenz der Modelle sind (Muthén, 1989; 1994). Um sich derjenigen Faktorenstruktur anzunähern, die die beobachteten Daten am wenigsten schlecht reproduziert, können die Within- und Between-Messmodelle separat mittels konventioneller CFA geschätzt werden. Die so ermittelten Modellparameter können als Startwerte für die MCFA genutzt werden. Hox (2002) und Stapleton (2006) liefern weitere, sehr ähnliche Vorgehensweisen.

Wie bereits erwähnt wird Σ_W durch S_{PW} konsistent und verlässlich geschätzt. Wenn das Konstrukt von Interesse und die zu überprüfende Faktorenstruktur in der Within-Ebene liegt und die Between-Cluster Abhängigkeiten nur ein Ärgernis darstellen, kann auf eine MCFA verzichtet werden und auf Grundlage von S_{PW} mittels klassischer CFA eine Faktorenstruktur sinnvoll falsifiziert werden (Muthén, 1994; Stapleton, Yang & Hancock, 2016). In diesem Fall sollte bei hohen ICCs eine konservativere Korrektur der Stichprobengröße vorgenommen werden als $N - G$. Denn wie erläutert, sollte mit steigenden ICCs die effektive Stichprobengröße geringer werden.

Eine hierarchische Datenstruktur mit Clusterabhängigkeit ist allerdings nicht nur ein Ärgernis, das es mit geeigneten Methoden zu berücksichtigen gilt, sondern liefert die Möglichkeit, die Faktorenstrukturen der jeweiligen Ebenen gleichzeitig zu modellieren und somit eine tiefgreifendere Erkundung des untersuchten Konstrukts.

Auswertung

Die GPS-Daten wurden gemeinsam mit den ES-Daten überprüft, um die Studienfahrten zu validieren. Es zeigte sich eine Variabilität bezüglich des tatsächlichen Beginns der Fahrten sowie der Distanz und Zeitspanne zwischen dem Beginn der Aufzeichnung und dem Erreichen der ersten Haltestelle. Zudem wurde die Station Rathaus häufig übersprungen. Aufgrund der lückenhaften GPS-Daten ist die Vergleichbarkeit der Fahrten eingeschränkt. Trotz dieser Variabilität wurden diese Fahrten in die Faktorenanalysen einbezogen, da sie die Validität der STS nicht beeinträchtigen. Zusätzlich zu den drei ausgeschlossenen Probanden und 13 kurzen, abgebrochenen Fahrten, wurden drei weitere Fahrten ausgeschlossen, da die Datenerfassung erst Stunden nach dem Ende der Fahrt abgeschlossen wurde und die STS die Bewertung zum Abschluss der Reise erfassen soll. Die Fahrdauer wurde als zeitliche Differenz zwischen Beginn und Ende der Aufzeichnung ermittelt.

Aufgrund anfänglicher Konvergenzproblemen der MCFA wurden klassische CFAs unter Verwendung der gepoolten Within- und angepassten Between Kovarianzmatrizen berechnet, um die Probleme des Messmodells zu identifizieren.

Sämtliche Datenaufbereitung und statistische Analysen wurden mit *R* (Version 4.3.1; *R* Core Team, 2024) durchgeführt. Die Faktorenanalysen wurden mit dem *R*-Paket *lavaan* (Version 0.6-13; Rosseel, 2024) durchgeführt. Zur Berechnung der benötigten Kovarianzmatrizen, des Skalierungsfaktors *c* und der ICCs wurde die Funktion von Huang (2017) verwendet, auf die die *lavaan*-Funktion basiert. Einschränkungen der MCFA-Funktion in *lavaan* im Vergleich zu *Mplus* (Muthén & Muthén, 2024) bestehen darin, dass nur kontinuierliche Variablen modelliert werden können und der weniger datenintensive Muthén-Maximum-Likelihood-Schätzer (Muthén, 1989) nicht implementiert ist.

Ergebnisse

Deskriptiv

Aus den insgesamt 387 analysierten Fahrten der 42 Probanden fanden 286 Studienfahrten auf dem vorgesehenen Weg mit einer durchschnittlichen Reisedauer von 57.1 Minuten ($SD = 17.4$) statt. Für 134 dieser Fahrten wurde der Bus als Hauptverkehrsmittel auf der Strecke zwischen Braunschweig und Wolfenbüttel anhand der GPS- und ES-Daten identifiziert ($M = 62.3$ Minuten, $SD = 17.9$) und für 149 Fahrten der Zug ($M = 52.8$ Minuten, $SD = 15.8$). 46 Fahrten wurden als Pendlerfahrten identifiziert, die über die Studienstrecke hinausgingen ($M = 91.6$ Minuten, $SD = 37.9$) und 33 davon mit dem Zug ($M = 90.6$ Minuten, $SD = 31.8$). 45 Fahrten waren Hin- und Rückfahrten zum oder vom DLR ($M = 39.8$ Minuten, $SD = 27$), 33 davon mit dem Bus ($M = 41.7$ Minuten, $SD = 22.3$). 10 Fahrten ließen sich nicht zuordnen ($M = 61.9$ Minuten, $SD = 37.3$), wurden jedoch als tatsächliche Fahrten identifiziert und zur Faktorenanalyse genutzt. Tabelle 2 beschreibt die STS erhoben zum Abschluss der 387 Fahrten ohne Berücksichtigung der Personenabhängigkeiten.

Tabelle 2*Korrelationen und Deskriptive Statistik der STS Items (N = 387)*

# Item	1	2	3	4	5	6	7	8	9
Korrelationen									
1 gehetzt - entspannt									
2 besorgt – zuversichtlich	.66								
3 gestresst - ruhig	.62	.56							
4 müde - wachsam	.32	.38	.34						
5 gelangweilt - enthusiastisch	.38	.43	.34	.53					
6 genervt - angeregt	.52	.48	.54	.45	.56				
7 Die Reise war die Schlechteste – Beste, die ich mir vorstellen kann	.51	.50	.43	.26	.46	.47			
8 Die Reise war von niedriger – hoher Qualität	.46	.45	.38	.30	.51	.49	.78		
9 Die Reise hat schlecht – reibungslos funktioniert	.44	.42	.39	.19	.35	.42	.62	.59	
Mittelwert	5.24	5.18	4.74	4.11	4.29	4.25	4.91	4.97	5.44
Standardabweichung	1.28	1.29	1.53	1.54	1.28	1.30	1.17	1.23	1.43
Schiefe	-.64	-.51	-.26	.20	.20	.00	-.09	-.15	-.62
Kurtosis	.19	.04	-.76	-.56	-.04	.06	-.31	-.23	-.56
ICC	.29	.35	.32	.36	.34	.56	.38	.39	.24

Bemerkungen. N = 387

Die hohen ICCs (vgl. Koch, 1983) der Indikatoren in Tabelle 2 indizieren starke Abhängigkeiten der einzelnen Bewertungen über die bewertenden Personen, die in weiteren Analysen berücksichtigt werden müssen. Die pooled Within- und angepasste Between-Korrelationsmatrizen sind in Tabelle 3 dargestellt und zeigen die bereinigten Interkorrelationen nach Berücksichtigung der Personenabhängigkeit auf.

Tabelle 3*Pooled Within-Korrelationsmatrix S_{PW} und angepasste Between-Korrelationsmatrix S_B*

# Item - Faktor	1	2	3	4	5	6	7	8	9
Pooled-Within									
Korrelationsmatrix									
1 - PD									
2 - PD	.52								
3 - PD	.51	.41							
4 - PA	.13	.21	.21						
5 - PA	.24	.29	.25	.45					
6 - PA	.40	.31	.43	.28	.38				
7 - CE	.44	.40	.36	.17	.37	.46			
8 - CE	.35	.29	.27	.15	.41	.49	.67		
9 - CE	.35	.26	.26	.14	.31	.41	.61	.60	
Angepasste-Between									
Korrelationsmatrix									
1 - PD									
2 - PD	.96								
3 - PD	.88	.89							
4 - PA	.74	.68	.60						
5 - PA	.71	.69	.52	.68					
6 - PA	.80	.80	.77	.75	.92				
7 - CE	.65	.70	.57	.41	.63	.48			
8 - CE	.69	.74	.59	.54	.69	.50	.96		
9 - CE	.71	.84	.74	.33	.46	.44	.67	.59	

Die unstandardisierten Kovarianzmatrizen S_{PW} und $c S_B$, auf deren Grundlage die MCFA geschätzt wird, sind im Anhang A dargestellt. Mit ihnen können die nachfolgenden Analysen selbst durchgeführt werden, um gegebenenfalls weitere Messmodelle zu testen. Aufgrund anfänglicher Konvergenzprobleme

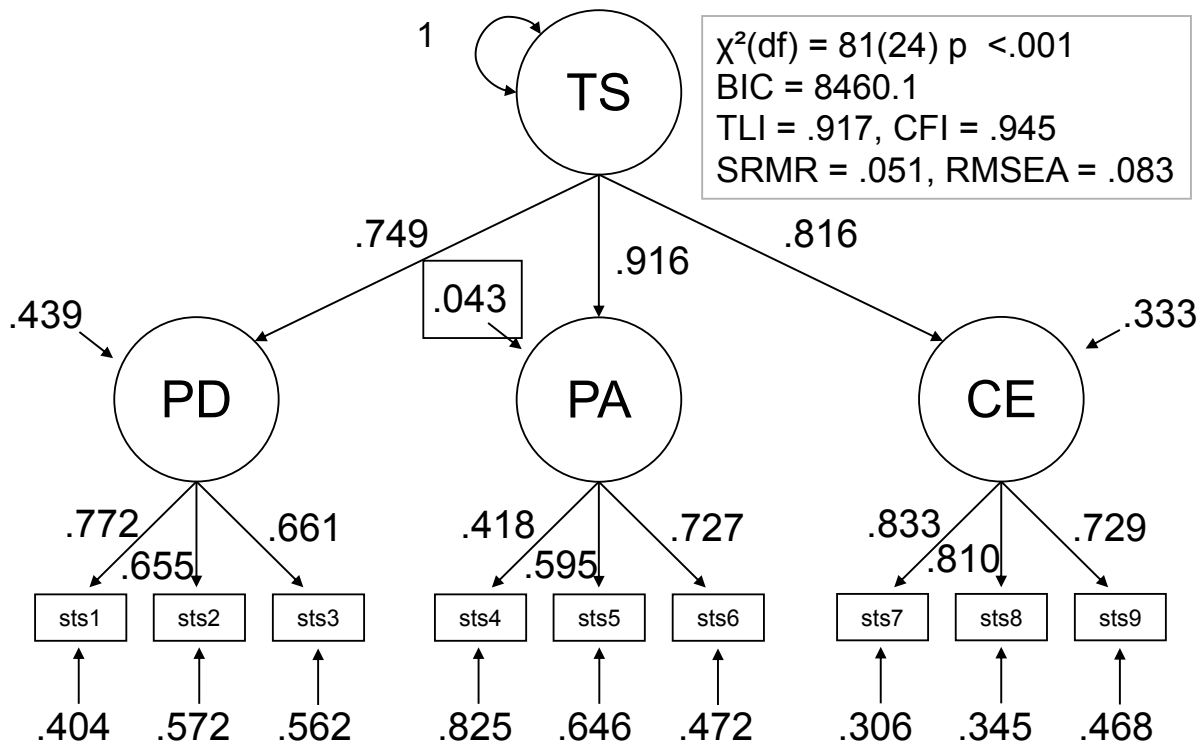
der MCFA, wurden die klassischen CFAs auf Grundlage dieser Kovarianzmatrizen zur Modellfindung durchgeführt und detailliert beschrieben, um die Entscheidungen der Modellmodifikationen nachvollziehbar zu beschreiben.

Die Reise zum Mehrebenenmessmodell

Abbildungen 3 und 4 zeigen die wesentlichen Within-Messmodelle und deren Modellpassungen.. Alle Modelle wurden mittels Maximum-Likelihood-Algorithmus (Satorra & Bentler, 1994) geschätzt. Referenzindikatoren skalieren die latenten Variablen. Die Within Modelle wurden mit $N - G = 345$ Beobachtungen geschätzt.

Das Within-Modell sollte der angenommenen hierarchischen Struktur der STS folgen, sofern Zufriedenheit als situationales Konstrukt betrachtet wird. Das in Abbildung 3 abgebildete Messmodell weist bereits keine zu schlechte globale Modellpassung auf. Im Vergleich dazu führt ein 2-Faktorielles Modell mit einer affektiven Bewertungsdimension, auf der die Items 1-6 laden und einer kognitiven Bewertungsdimension, auf der die Items 7-9 laden, zu einer schlechteren Modellpassung ($\chi^2 (df) = 131.59 (26) p < .001, TLI = .859, CFI = .898, SRMR = .067, RMSEA = .108, BIC = 8499$). Ein 1-Dimensionales Messmodell weist eine noch schlechtere globalen Modellpassung auf ($\chi^2 (df) = 221.24 (27) p < .001, TLI = .750, CFI = .812, SRMR = .086, RMSEA = .144, BIC = 8582.8$).

Die in Abbildung 3 dargestellten globalen Fitstatistiken, wie der *RMSEA* von .083, indizieren einerseits bereits eine hinreichende Passung des Messmodells zu den beobachteten Daten (Hu & Bentler, 1999) andererseits implizieren sie Verbesserungspotenzial für das Modell, die durch Betrachtung der lokalen Fitstatistiken ersichtlich werden können. Im hierarchischen Messmodell mit allen Items (Abbildung 3) sind alle geschätzten Parameter signifikant ($p < .001$), mit Ausnahme der Disturbance der PA-Dimension ($p = .134$). Bis auf die zu geringen Faktorladungen der Items 4 und 5 weisen die Items zufriedenstellende Ladungen auf ihren jeweiligen Faktoren auf (z.B. Gäde & Schermelleh-Engel, in Moosbrugger & Kelava, 2020, S. 648-651).

Abbildung 3*Hierarchisches Within-Modell mit allen 9 Items der STS*

Bemerkungen. TS = Reisezufriedenheit, PD = Positive Deaktivierung, PA = Positive Aktivierung, CE = Kognitive Evaluation. Die markierte Disturbance von PA ist nicht signifikant ($p = .134$) alle anderen geschätzten Parameter sind signifikant mit $p < .001$.

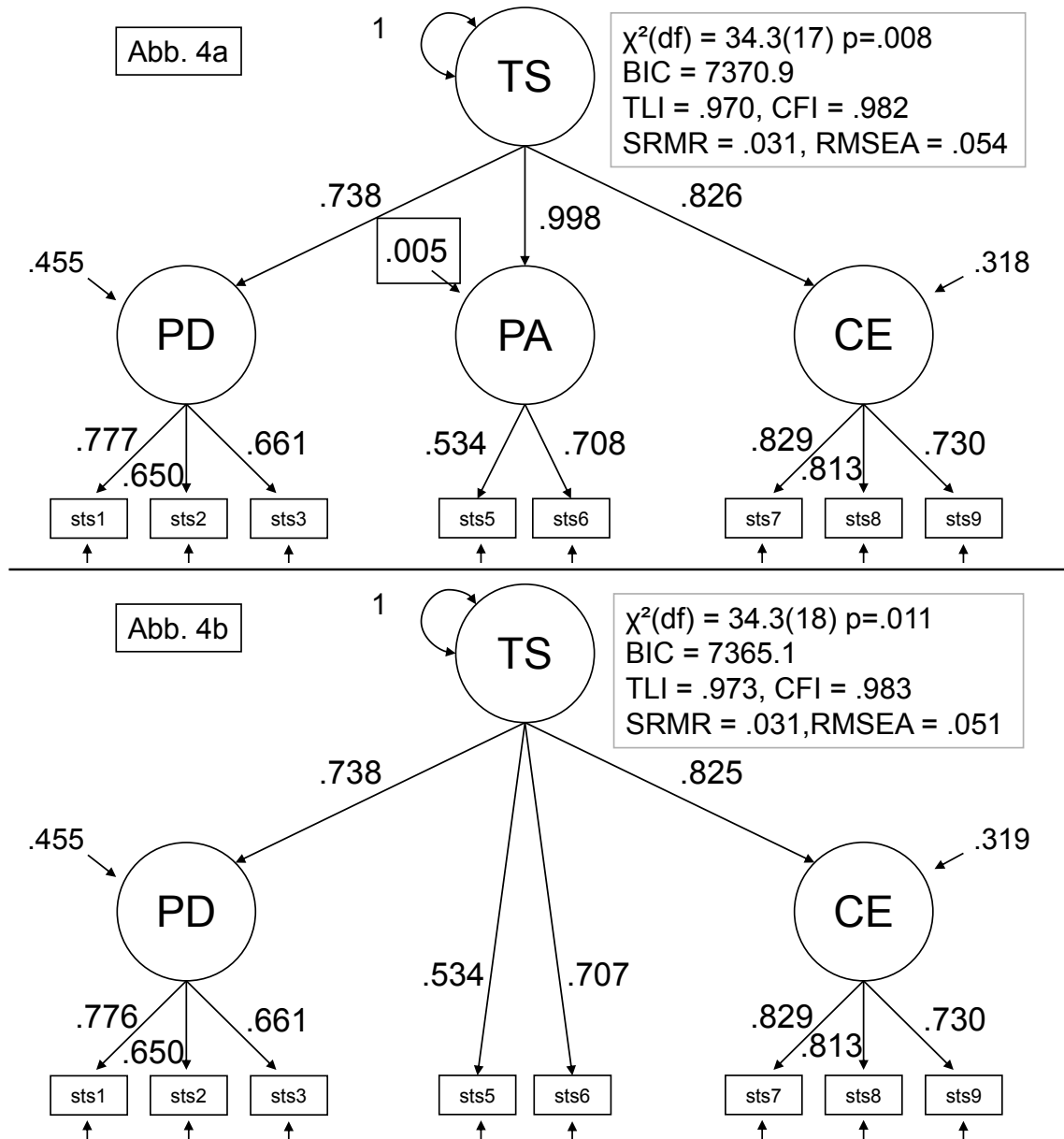
Der Grund der schwachen lokalen Modellpassung von Item 4 liegt vermutlich in der Itemformulierung. Die übrigen Items teilen eine gemeinsame Skalierung, bei der extreme Bewertungen an den Skalenenden und neutrale Bewertungen im mittleren Bereich liegen. Durch die Formulierung der Anker von Item 4 („müde“ und „wachsam“) liegt die Bewertung der höchsten Valenz eher im mittleren Skalenbereich, falls Valenz im Gegensatz zu Arousal überhaupt in diesem Item abgebildet ist.

Das Messmodell ohne Item 4 (Abbildung 4a) weist eine verbesserte globale Modellpassung auf. Die nahezu perfekte Ladung von PA auf TS (.998) indiziert, dass diese Dimensionen praktisch identisch sind. Daher wird auf die Modellierung von PA verzichtet und die Items 5 und 6 laden direkt auf TS

(Abbildung 4b). Da das schwach ladende Item 5 nicht erforderlich ist, um einen Faktor zu modellieren, kann auf dieses, auch zugunsten des folgenden Between-Modells, verzichtet werden.

Abbildung 4

Hierarchisches Within-Messmodell der STS ohne Item 4 (müde-wachsam)



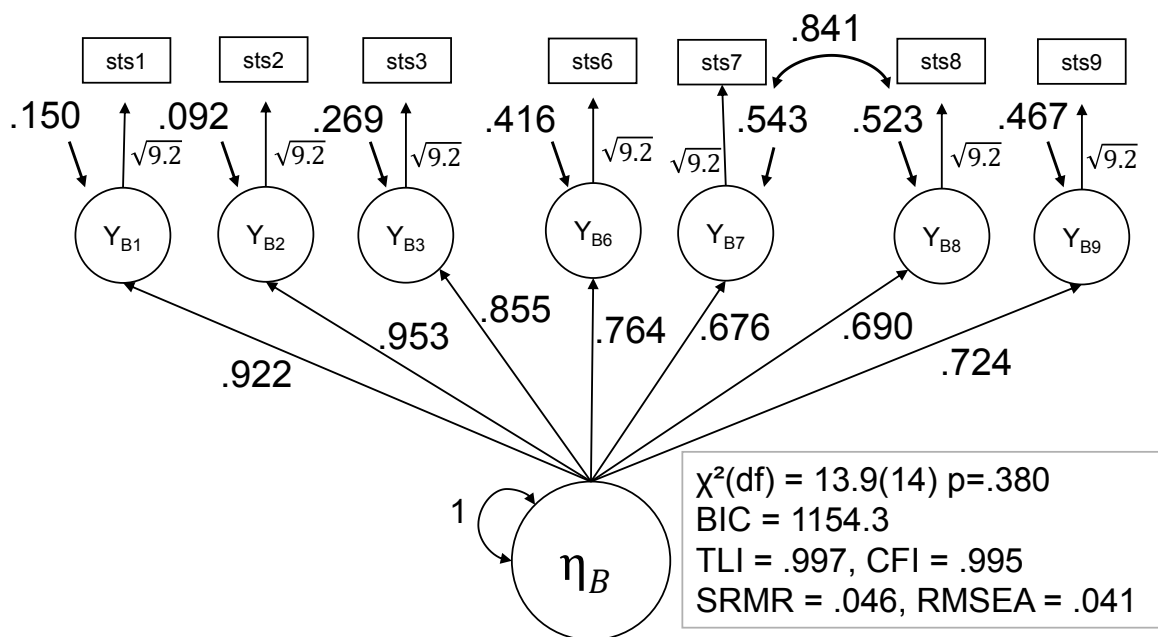
Bemerkungen. Disturbance von PA ist nicht signifikant ($p = .969$). Durch Einsparung des Faktors 1.

Ordnung PA wird das Modell sparsamer, verschlechtert sich nicht und alle lokalen Fitstatistiken sind signifikant ($p < .001$).

Mit den sieben verbliebenen Items wurde ein Between-Modell mit $G = 42$ Beobachtungen geschätzt. Vermutlich wegen der kleinen Stichprobengröße konvergiert nur ein sparsames 1-Faktoren Modell sinnvoll, allerdings mit schlechter globaler Modellpassung ($\chi^2 (df) = 63 (14) p < .001$, $TLI = .722$, $CFI = .814$, $SRMR = .089$, $RMSEA = .291$, $BIC = 1200.6$). Durch das Zulassen kovariierender Residuen zwischen den Items 7 („Die Reise war die schlechteste ... beste, die ich mir vorstellen kann“) und 8 („Die Reise war von niedriger ... hoher Qualität“) verbessert sich die Modellpassung deutlich (Abbildung 5; $\chi^2_{diff} (1) = 50.05$, $p < .001$).

Abbildung 5

1-Faktorielles Between-Messmodell mit kovariierenden Residuen der Items 7 und 8



Bemerkungen. Der Skalierungsfaktor $c = 9.2$ entspricht der durchschnittlichen Clustergröße (der durchschnittlichen Anzahl an Fahrten pro Person)

Diese Kovarianz könnte darin begründet sein, dass Items 7 und 8, im Gegensatz zu den übrigen Items, einen Vergleich der eben erlebten Reise mit einem mentalen Modell vergangener erlebter Reisen erfordern. Das mentale Modell einer typischen Reise sollte innerhalb einer Person über die Fahrten

hinweg eher konstant bleiben und zwischen verschiedenen Personen variieren. Daher tritt diese Residualkovarianz auf der Between- und nicht auf der Within-Ebene auf.

Mehrebenenmessmodell (MMM)

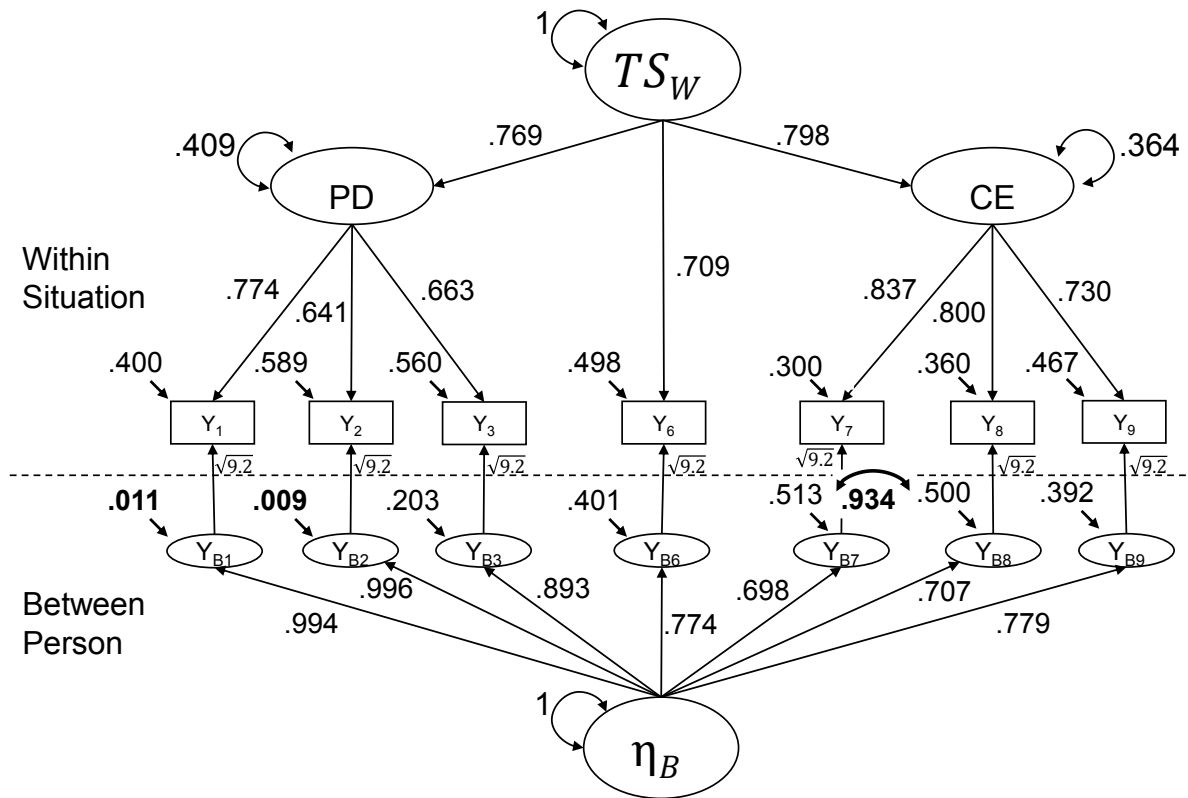
Die modifizierten Within- und Between Modelle liefern die Struktur für das in Abbildung 6 dargestellte und mittels MCFA analysierte MMM. Die auf Basis von $N = 387$ Beobachtungen und $G = 42$ Personen mittels robuster Maximum-Likelihood Schätzung (Yuan & Bentler, 1999) ermittelten Modellparameter indizieren eine gute Passung des Modells zu den beobachteten Daten ($\chi^2(df) = 30.81(27)$ $p = .279$ (Yuan-Bentler), $_{Robust}TLI = .992$, $_{Robust}CFI = .995$, $SRMR_{Within} = .028$, $SRMR_{Between} = .052$, $_{Robust}RMSEA = .021$, $BIC = 7662.0$). Alle Modellparameterschätzungen sind signifikant mit $p < .05$.

Die sehr hohen Faktorladungen der Items 1 und 2 auf den Between-Faktor bringen minimale Messfehler mit sich. Bei MCFA Schätzungen treten häufig negative Residuen auf der Between-Seite auf. Wenn die Standardfehler der betroffenen Items darauf hinweisen, dass die Fehlervarianzen tatsächlich sehr nah bei null liegen, können diese auf null oder einen sehr kleinen positiven Wert fixiert werden, um eine sinnvolle Modellkonvergenz zu ermöglichen (Hox, 2002; Muthén & Asparouhov, 2011; Kolenikov & Bollen, 2012). Deshalb wurden die Between-Residuen der Items 1 und 2 auf .005 fixiert.

Der nicht-signifikante χ^2 -Test indiziert, dass die Nullhypothese der Modelltestung, dass kein Unterschied zwischen der modellimplizierten Kovarianzmatrix und der empirischen Kovarianzmatrix besteht, nicht verworfen werden sollte. Das bedeutet, dass das MMM die Zusammenhänge der beobachteten Daten optimal abbildet. Dementsprechend bedarf es keiner Testung komplizierterer Modelle.

Abbildung 6

MCFA mit den verbliebenen 7 Items der STS



Bemerkungen. Fett markierte Parameter wurden fixiert. Die geschätzten Intercepts wurden aus Darstellungsgründen ausgelassen ($v_1 = 5.23$, $v_2 = 5.16$, $v_3 = 4.73$, $v_6 = 4.2$, $v_7 = 4.88$, $v_8 = 4.95$, $v_9 = 5.44$)

Im Vergleich zum hierarchischen Within-Modell reproduziert ein sparsameres 1-Faktorielles Within-Modell die beobachteten Daten schlechter ($\chi^2(df) = 114.06 (29) p < .001$ (Yuan-Bentler), $Robust TLI = .847$, $Robust CFI = .895$, $SRMR_{Within} = .086$, $Robust RMSEA = .102$, $BIC = 7768.1$). Ein 2-Faktorielles Within-Modell, wie im vorherigen Abschnitt beschrieben, weist eine nur geringfügig schlechtere Modellpassung auf ($\chi^2(df) = 50.74 (28) p = .005$ (Yuan-Bentler), $Robust TLI = .962$, $Robust CFI = .975$, $SRMR_{Within} = .053$, $Robust RMSEA = .051$, $BIC = 7680.4$). Die Faktoren des 2-Faktoriellen Modells korrelieren relativ hoch miteinander ($r = .69$, $p < .001$). Diese hohe Korrelation deutet möglicherweise auf eine hierarchische Lösung mit Faktor 2. Ordnung hin, die mit lediglich 2 Faktoren 1. Ordnung nicht identifiziert ist. Gemäß des

Sparsamkeitsprinzips sollte das 2-Faktorielle Modell nicht zwingend verworfen werden. Dennoch wären nachfolgende Regressionsmodelle mit einem einzelnen Faktor effizienter, insofern dieser die gesamte Bewertung adäquat repräsentiert.

Die Berechnung der Reliabilitäten der Within- und Between- Modelle erfolgt separat (Geldhof, Preacher & Zyphur, 2014; Muthén, 1994). Der Varianzanteil des Between-Faktors an der gesamten modellierten Between-Varianz beträgt $\omega_b = .897$ (Bollen, 1980). Der Varianzanteil der Within-Faktoren TS, PD und CE an der gesamten modellierten Within-Varianz (Omega_Total) beträgt $\omega_{TW} = .871$. Der Anteil an der modellierten Within-Varianz, der ausschließlich auf den hierarchischen Within-Faktor TS zurückzuführen ist, beträgt $\omega_{HW} = .716$ (vgl. Reise, Bonifay & Haviland, 2013). Details der Reliabilitätsberechnung sind im Anhang B dargestellt.

Die Varianz der Summenwerte der genutzten 7 Items der STS lässt sich zu 60 % durch den hierarchischen Within-Faktor erklären ($F_{1,385} = 583.6, p < .001, R^2 = .601$) und zu 96% durch eine additive Linearkombination des hierarchischen Within- und des Between-Faktors ($F_{2,384} = 4577, p < .001, R^2 = .960$)

Interpretation der Faktorenstruktur der STS

Eine hierarchische Struktur der Reisezufriedenheit erweist sich als die beste Lösung zur Reproduktion der beobachteten Daten, trotz der Reduktion auf sieben verbliebene Items. Die aus der Literatur (Friman et al., 2013; Singleton, 2019) bekannte Faktorenstruktur findet sich in der Within-Ebene wieder und bestätigt dessen situative Natur. Wichtig ist, dass ein 1-Faktorielles Modell die Daten schlechter reproduziert als das hierarchische oder 2-Faktorielle Modell. Für die Modellierung der Reisezufriedenheit als situatives Konstrukt ist es entscheidend, dass die Items zur kognitiven Bewertung der Fahrt nicht auf denselben Faktor laden wie die kognitiven Bewertungen des erlebten Affekts.

Im Gegensatz dazu ist die Between-Ebene des Modells eindimensional. Ein hoher Between-Faktorscore bedeutet, dass eine Person über alle Fahrten hinweg tendenziell positiver bewertet, unabhängig davon, ob das Item die kognitive Bewertung der Fahrt oder des erlebten Affekts erfasst. Der

Between-Faktor repräsentiert also eine allgemeine **Neigung zu positiven oder negativen Bewertungen**. Er repräsentiert keine stabile Reisezufriedenheit als Pendant zur situativen Reisezufriedenheit. Für letztere Interpretation wäre mindestens eine faktorielle Invarianz zwischen den Ebenen erforderlich. Einige Autoren bestehen zusätzlich auf eine Invarianz der Faktorladungen über die Ebenen für solch eine Interpretation (Stapleton et al, 2016; Jak, 2019). Die Neigung zu positiven Bewertungen ist außerdem unabhängig von einem mentalen Modell einer typischen oder optimalen Reise einer Person, repräsentiert durch die zugelassenen kovariierenden Residuen der Between-Seite. Ein solches Modell könnte allerdings eine zentrale Facette einer stabilen Reisezufriedenheit sein.

Die Generalisierbarkeit der Neigung zu positiven Bewertungen ist anhand der vorliegenden Daten nicht überprüfbar, da der Personenfaktor ausschließlich durch abschließende Bewertungen multimodaler ÖPNV-Fahrten erfasst wurde.

Explorative Analysen

Um ein besseres Verständnis für die Bedeutung der Faktoren zu erlangen, wurden explorative Analysen mit den Factorscores in Verbindung mit weiteren in der Studie erhobenen Variablen durchgeführt. Tabelle 4 stellt den Zusammenhang verschiedener Ereignisse, die während der Fahrten berichtet wurden, mit den Factorscores der latenten Variablen des Within-Messmodells dar.

Einige der Ereignisse wie Verspätungen, Störende Menschen und Unwohlsein, ausgelöst z.B. durch Kälte, zeigen einen geringen, signifikanten, negativen Zusammenhang mit den latenten Faktoren der Reisezufriedenheit. Die gleichen Korrelationen der Ereignisse mit der latenten kognitiven Evaluation sind immer mindestens so stark, eher stärker als mit dem hierarchischen Faktor der Reisezufriedenheit, der auch die positive Deaktivierung erklärt. Die Korrelationen der Ereignisse mit der latenten positiven Deaktivierung sind dagegen höchstens so stark wie die latente Bewertung der Reise oder nicht signifikant.

Tabelle 4

Punktbiserial Korrelationen zwischen den Faktorscores des Within-Messmodells und während der Fahrten berichteten Ereignissen

Ereignis	N (%) ^a	TSw	CE	PD
1 Verspätung / Anschluss verpasst / Beeilung nötig	127	-.14**	-.18***	-.12*
2 Störende Menschen in der Umgebung/ Überfülltes Verkehrsmittel	131	-.12**	-.14**	-.08
3 Fahrverhalten des Fahrers	30	-.05	-.06	-.05
4 Infrastruktur (z.B. keine Überdachung/ keine oder nicht genügend Informationen / nicht barrierefrei / unebene Straße	38	-.05	-.05	-.03
5 Positive soziale Situation / Interaktion	84	.03	.03	.04
6 Mediale Unterhaltung	104	.07	.07	.05
7 Ziel oder Anschluss erreicht / Verkehrsmittel pünktlich	182	.09	.10	.03
8 Unwohlsein (Langeweile, Ungeduld, schwierige Gedanken, Hunger, Müdigkeit, Frieren, Nässe, etc)	124	-.12*	-.12*	-.12*
9 Verkehrsmittel ist komfortabel	187	.13*	.13**	.09
10 Schöne Umgebung	98	.09	.07	.07
11 Private / berufliche Erledigungen	52	-.01	-.01	-.05
12 Sonstiges	147	-.07	-.07	-.06

Bemerkungen. TSw = Reisezufriedenheit, CE = Kognitive Evaluation, PD = Positive Deaktivierung

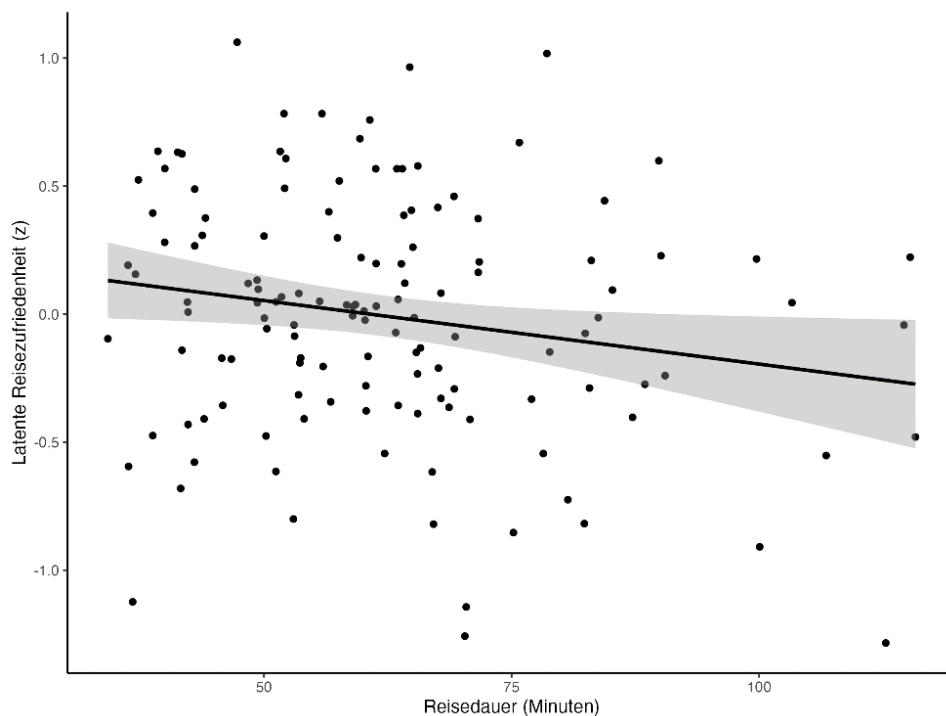
^aN (%) = Häufigkeit der Fahrten, in denen das Event mindestens 1 Mal berichtet wurde.

* $p < .05$. ** $p < .01$. *** $p < .001$. $df(N-2) = 385$.

Eine lineare Regression mit der errechneten Fahrtdauer als Prädiktor zeigt keinen linearen Zusammenhang mit der latenten Reisezufriedenheit ($F_{1,385} = .22, p = .639$), der latenten kognitiven Evaluation ($F_{1,385} = 1.45, p = .229$) oder der latenten positiven Deaktivierung ($F_{1,385} = .003, p = .954$). Wenn nur die 134 Fahrten berücksichtigt werden, in denen die Studienstrecke wie vorgesehen und mit dem Bus als Hauptverkehrsmittel zwischen Braunschweig und Wolfenbüttel die Hauptstrecke absolviert wurden, zeigt sich ein geringer signifikanter linearer Zusammenhang, der in Abbildung 7 beschrieben ist ($F_{1,132} = 4.881, p = .029, \beta = -.005, R^2 = .028$). Interessanterweise zeigt sich ein geringerer, nach der klassischen Fisher-Signifikanzgrenze von $p = .05$ (Fisher, 1925) nicht mehr signifikanter Zusammenhang für die latente kognitive Evaluation ($F_{1,132} = 3.698, p = .057, \beta = -.006, R^2 = .020$) und positive Deaktivierung ($F_{1,132} = 3.381, p = .068, \beta = -.006, R^2 = .018$).

Abbildung 7

Linearer Zusammenhang zwischen der Reisedauer und der latenten Reisezufriedenheit für Busfahrten auf der Studienstrecke

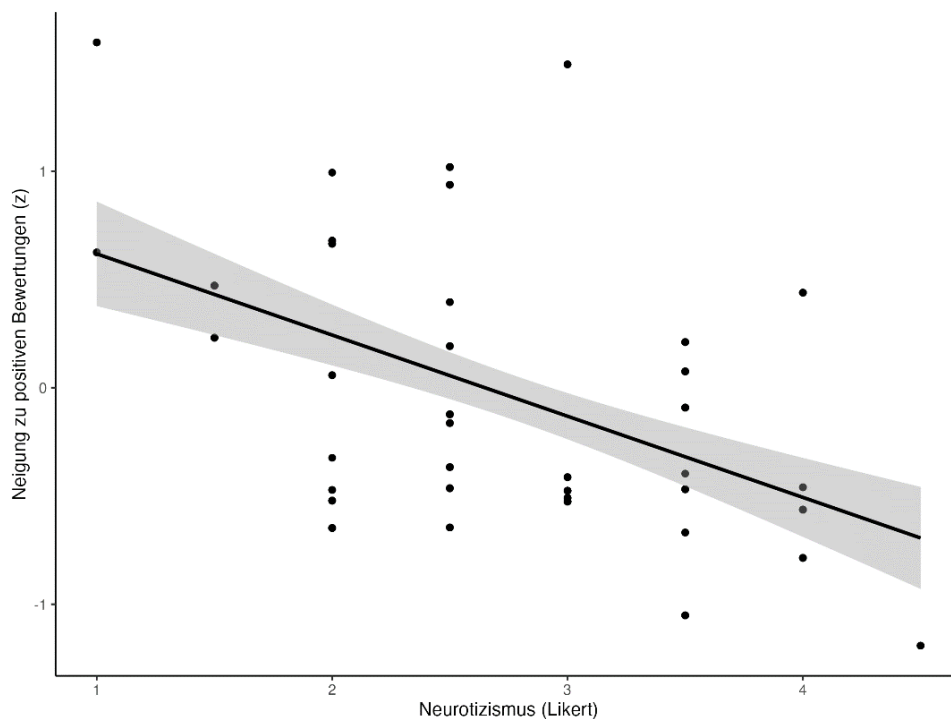


Bemerkungen. N = 134

Zur Einordnung der Neigung zu positiven Bewertungen wurden Spearman-Rangkorrelationen berechnet zwischen den Factorscores und den Mittelwerten der Big-Five-Persönlichkeitsdimensionen, erhoben mit dem BFI-10 (Rammstedt et al., 2013). Die Neigung zur positiven Bewertung korreliert signifikant mit Neurotizismus ($\rho = -.41, p = .014$), jedoch nicht mit Extraversion ($\rho = .24, p = .116$), Offenheit ($\rho = -.05, p = .716$), Verträglichkeit ($\rho = .18, p = .266$) oder Gewissenhaftigkeit ($\rho = .17, p = .273$). Mit steigendem Neurotizismuswert einer Person, tendiert sie eher zu negativen Bewertungen. Abbildung 8 beschreibt diesen linearen Zusammenhang ($F_{1,40} = 8.926, \beta = -.358, p = .005, R^2 = .016$).

Abbildung 8

Linearer Zusammenhang zwischen der Neigung zu positiven Bewertungen und Neurotizismus



Bemerkungen. N = 42

Reisezufriedenheit und Verkehrsmittelpräferenz

Im Abschlussfragebogen mussten die Probanden sich entscheiden, ob sie die Strecke zwischen Braunschweig und Wolfenbüttel zukünftig bevorzugt mit dem Bus oder mit dem Zug zurücklegten.

In den insgesamt 40 ausgefüllten Abschlussfragebögen äußerten 12 Personen eine Präferenz für den Bus und 28 Personen für den Zug.

T-Tests für unabhängige Stichproben zeigten keine signifikanten Unterschiede in der latenten Reisezufriedenheit ($t(280.94) = -.597, p = .551$) oder der latenten kognitiven Bewertung ($t(280.77) = -.704, p = .482$) zwischen Bus und Zug als Hauptverkehrsmittel auf der Strecke zwischen Braunschweig und Wolfenbüttel an. Innerhalb der 12 Pro-Bus-Probanden wurden ebenfalls keine Unterschiede in der latenten Reisezufriedenheit ($t(73.95) = -.386, p = .701$) oder kognitiven Bewertung ($t(73.23) = -.271, p = .787$) zwischen den Hauptverkehrsmitteln festgestellt. Ebenso zeigten die 28 Pro-Zug-Probanden keinen Unterschied hinsichtlich ihrer latenten Reisezufriedenheit ($t(192.89) = -.832, p = .406$) oder latenten kognitiven Bewertung ($t(192.76) = -.931, p = .3533$).

Die Präferenzen für ein Verkehrsmittel lassen sich demnach auf Grundlage der in dieser Studie erhobenen Daten nicht durch Unterschiede in den Zufriedenheitsbewertungen mit den Hauptverkehrsmitteln erklären. Nach der Verkehrsmittelpräferenz wurde im offenen Antwortformat nach einer Begründung für die Präferenz gefragt. Um zu verstehen, wieso mehr Personen den Zug auf dieser Strecke präferieren, sind in Tabelle 5 die Argumente für das bevorzugte Verkehrsmittel ihrer genannten Häufigkeit nach aufgelistet, die eine Zusammenfassung der offenen Befragung zum Abschluss der Studie darstellt.

Der häufigste genannte Grund für eine Präferenz für den Zug war die kürzere Fahrzeit, wobei in einem Fall als Pro-Argument für den Bus genannt wurde, dass durch die längere Fahrzeit mehr Zeit für Tätigkeiten bestehe (Komfort), was möglicherweise auch das Argument „kürzere Wartezeiten“ erklärt. Weitere Komfortaspekte des Busses wurden nicht genannt. Für den Zug war Komfort hingegen ein häufig

genanntes Argument, beispielsweise sei die Fahrt mit dem Zug weniger ruckelig und leiser, die Sitze seien bequemer und die Temperatur besser reguliert. Sechs Probanden verwiesen auf eine höhere Zuverlässigkeit des Zuges. Fünf Probanden präferierten hingegen die höhere Taktung des Busses, wodurch dieser resilienter gegen Ausfälle sei.

Tabelle 5

Argumente für das präferierte Verkehrsmittel zukünftiger Fahrten zwischen Wolfenbüttel und Braunschweig

Pro Zug (28)	Pro Bus (12)
Kürzere Fahrzeit (24)	Höhere Taktung (5)
Komfort (15)	Weniger Umstiege (5)
Mehr Sitzplätze (7)	Komfort (2)
Zuverlässiger (6)	Kürzere Wartezeiten (2)
Weniger Zwischenstopps (4)	Attraktivere Umgebung (2)
Unkomplizierter (2)	Nähe zum Wohnort (1)
Ausstattung (2) – Steckdosen (1) und Toilette (1)	Mehr soziale Interaktion (1)
Attraktivere Umgebung (1)	
„fahre sowieso lieber Zug“ (1)	

Bemerkungen. N = 40. 2 Abschlussfragebögen wurden nicht ausgefüllt. Die meisten Antworten enthielten mehrere Argumente.

Diskussion

Diese Arbeit untersucht die psychometrischen Eigenschaften der Satisfaction with Travel Scale, die zur Bewertung tatsächlich erlebter multimodaler ÖPNV-Fahrten eingesetzt wurde. Die unbalancierten Längsschnittdaten konnten erfolgreich mittels MCFA in *R Lavaan* (2024) faktorenanalytisch modelliert werden. Nach Ausschluss der Items 4 und 5 sowie der Berücksichtigung kovariierender Residuen der Items 7 und 8 auf der Between- Seite des Messmodells, reproduzierte das MMM die erhobenen Daten mit einem hierarchischen Modell mit der Reisezufriedenheit auf der Within-Seite und einem eindimensionalen Messmodell auf der Between-Seite, nahezu perfekt ($\chi^2(df) = 30.81(27) p=.279$). Der eindimensionale Between-Faktor wird als situationsunabhängige Neigung zu positiven oder negativen Bewertungen interpretiert.

Die mittlere Korrelation mit Neurotizismus - obwohl dieses Konstrukt, oder zumindest eine Facette dessen, nur mittels zweier Likert-Skalen Items erfasst und nicht um dessen Messfehler korrigiert wurde - unterstützt diese Interpretation. Costa und McCrae (1980) argumentieren, dass es glückliche und unglückliche Menschen gibt, eher als Konsequenz ihrer Persönlichkeit anstelle individueller Lebensumstände. Zufriedenheitsbewertungen sollten prinzipiell durch den Bewertungsgegenstand determiniert sein, sind aber ebenfalls durch die Persönlichkeit des Bewertenden gefärbt. Diese Arbeit zeigt exemplarisch am Konstrukt der Reisezufriedenheit, dass die Ursachen der Bewertung des Reiseerlebens sowohl in der erlebten Situation als auch der bewertenden Person liegen, die unabhängig von der Situation zu positiven oder negativen Einschätzungen neigt. Unterschiede zwischen den Reisen, ob in der äußeren Welt oder der inneren Welt eines Reisenden, erklären 60% der Varianz der durch die STS erhobenen Testwerte. Eine Linearkombination aus der voneinander unabhängig modellierten situativen Reisezufriedenheit und der persönlichen Neigung zu positiven oder negativen Bewertungen erklärt 96% der Varianz. Der Varianzanteil, der auf die bewertende Person zurückzuführen ist, sollte berücksichtigt werden.

Bisherige psychometrische Analysen berichten auch nach datengetriebenen Modellmodifikationen geringere Bewertungen der Modellgüte, Singleton (2019) lieferte die bisher beste Annäherung. Auch einmalige Zufriedenheitsmessungen ergeben sich aus einer Funktion aus dem Bewertungsgegenstand und der bewertenden Person, bloß ist der Einfluss der Person nicht vom Einfluss der Situation differenzierbar, genauso wie aus einer einmaligen Verhaltensbeobachtung keine Persönlichkeitseigenschaft abgeleitet werden kann (Mischel, 1969). Deshalb konnten bisherige psychometrische Untersuchungen der STS trotz datengetriebener Modellspezifikationen keine optimierte Modellgüte erreichen, weil sich das eindimensionale Between-Modell und das hierarchische Within-Modell überlagern und anhand von Querschnittsdaten nicht getrennt modelliert werden können. Die eindimensionale Neigung zu positiven oder negativen Bewertungen verwässert das hierarchische Messmodell der Reisezufriedenheit.

Die konfirmatorische Natur der MCFA ist aufgrund der datengeleiteten Modellmodifikationen aufgehoben. Dennoch ist es bemerkenswert, dass sich das hierarchische Within-Modell durchsetzt, obwohl nur sieben Items verblieben sind und aus der ursprünglichen Facette positive Aktivierung nur ein Item überlebte, das direkt auf dem Faktor 2. Ordnung lädt. Da ein 2-faktorielles Within-Modell nur geringfügig schlechter passt, ein eindimensionales dagegen deutlich schlechter, ist es wichtig zur Modellierung der STS beziehungsweise zur Modellierung der Reisezufriedenheit als situatives Konstrukt, dass die Items zur kognitiven Bewertung der Fahrt und des Affekts auf verschiedenen Faktoren laden.

Diese Faktoren 1. Ordnung werden als Facetten der Reisezufriedenheit interpretiert, da sie hoch miteinander korrelieren, auch in einem 2-Faktoriellen Modell. Allerdings rechtfertigt eine starke Korrelation latenter Variablen nicht zwangsläufig die Modellierung eines Faktors höherer Ordnung. Explorative Analysen zeigen geringe Zusammenhänge zwischen während der Fahrt berichteten Erlebnissen, wie Verspätungen oder störenden Mitmenschen und der latenten Reisezufriedenheit als

Faktor 2. Ordnung. Doch alle signifikanten Zusammenhänge sind gleich stark oder sogar stärker mit der latenten kognitiven Evaluation, die lediglich drei Items erklärt.

Dieses konsistente Bild zeichnet sich für die berichteten Ereignisse über alle Fahrten, allerdings nicht für die Dauer der Fahrten zwischen Braunschweig und Wolfenbüttel mit dem Bus als Hauptverkehrsmittel. Diese weisen einen verschwindend geringen signifikanten Zusammenhang zwischen der Fahrtdauer und der latenten Reisezufriedenheit, jedoch nicht mit der kognitiven Evaluation oder positiven Deaktivierung auf. Insgesamt spielt die Reisezeit keine zentrale Rolle für die latente Reisezufriedenheit und dessen Facetten, selbst bei Fahrten mit vergleichbaren Erwartungen an die Reisedauer. Dieser fehlende Zusammenhang überrascht, da mehr als die Hälfte der Probanden ihre Verkehrsmittelpräferenz mit einer kürzeren Fahrtdauer begründen. Ob die Reisezufriedenheit oder die Reisedauer zu ungenau erfasst wurde oder ein *Extension Neglect* vorliegt, der den Probanden nicht bewusst ist, bleibt offen. Die geringen Korrelationen der Ereignisse während der Reisen mit der Reisezufriedenheit sprechen für die Validität des Konstrukts, zumindest für die kognitive Facette. Allerdings untergraben die Items zur Bewertung des erlebten Affekts die Konstruktvalidität der STS und verwässern die Bewertung des Reiseerlebens.

Es erscheint zunächst plausibel, dass sich Reisezufriedenheit im erlebten Affekt widerspiegeln sollte. Eine Fahrt, die ein niedriges Arousal und eine hohe Valenz hervorruft, also entspannt und mit positiven Emotionen verbunden ist, dürfte als angenehmer wahrgenommen werden als eine, die hohe körperlicher Erregung und negative Gefühle auslöst. Sowohl die Appraisal-Theorie (Scherer, 2001; 2019) als auch Barretts konstruktivistische Perspektive (Barrett, 2017) unterstützen die Idee, dass Affekte kontextabhängige Konstruktionen sind. Allerdings stellt sich die Frage, ob die in der STS verwendeten Items tatsächlich den erlebten Affekt erfassen oder vielmehr kognitive Konstruktionen widerspiegeln, bei denen die Adjektive auf generelle positive oder negative Eigenschaften der Fahrt, statt auf das emotionale Erleben zum Zeitpunkt der Erhebung, zurückgeführt werden. Selbst wenn der Einfluss der Reise auf den

Affekt bloß geringer ist als auf die kognitive Evaluation, sind in beiden Szenarien sechs Items zur Bewertung des Affekts aus insgesamt neun Items eine starke Überrepräsentation des unklar definierten Konstrukts und bedarf einer Anpassung der Skala.

Die Darstellung der Modellentwicklung in dieser Arbeit mag zunächst ungewöhnlich detailliert erscheinen, doch sind es solche psychometrischen Details, die entscheidend sein können, um die Weiterentwicklung von Skalen und das Verständnis der gemessenen Konstrukte voranzutreiben. Wie Wilson (2023) betont, ist die Skalenentwicklung ein iterativer Prozess, der in einem Kreislauf des Erkenntnisgewinns stattfindet. Die Erkenntnisse dieser Arbeit implizieren, dass zukünftige Skalen zur Erfassung der Reisezufriedenheit entweder auf Items zur Erfassung des Affekts verzichten sollten oder die Anzahl im Verhältnis zur kognitiven Evaluation der Fahrt deutlich reduzieren. Da ein Faktor, der lediglich durch 3 Items modelliert wird, keine sparsame Lösung darstellt und zudem keine umfassende Erfassung eines Konstrukts ermöglicht, bedarf es mehr kognitiver Indikatoren.

Die Kovarianz der Residuen der Items 7 und 8 auf der Between-Seite des MMM bildet eine Einschätzung der Reise im Vergleich zu früheren Erfahrungen ab, die durch die Frage nach einem reibungslosen Verlauf nicht erfasst wird. Diese Kovarianz spricht zunächst gegen die Verwendung dieser Items, da sie die Modellannahmen verletzt. Statt diese Items zu verwerfen, sollten mehr Items hinzugefügt werden, um diese spezifischen Facetten der kognitiven Bewertung faktorenanalytisch modellieren zu können. Eine psychometrische Analyse der Satisfaction with Life Scale (Diener, Emmons, Larsen & Griffin, 1985) unterstützt diese Überlegung. Die Skala erfasst die Zufriedenheit mit dem eigenen Leben mittels fünf Likert-Items, von denen zwei einen Rückblick auf das gesamte Leben erfordern, während die anderen drei die aktuelle Zufriedenheit erfragen. Eine Residualkovarianz der zwei rückblickenden Items verbessert die Modellpassung signifikant ($\chi^2_{diff}(1) = 10.21, p = .0014$ im Datensatz von Wetzel & Frick, 2020 mit $n = 763$; Janke & Glöckner-Rist, 2014). Eine klare Unterscheidung der Bezugspunkte der kognitiven Bewertung bzw. dessen Referenzrahmen sollte in der Itemformulierung berücksichtigt werden. Nicht nur die STS,

sondern Skalen zur Zufriedenheitserfassung im Allgemeinen können davon profitieren, möglichst viele klar formulierte Items zu verwenden, um die jeweiligen Referenzrahmen der kognitiven Bewertung als eigene Facetten zu berücksichtigen und somit ein Zufriedenheitskonstrukt reliabler und valider zu erfassen. Eine bloße Auflistung von Serviceaspekten sollte jedoch vermieden werden, um die reflektive Natur des Messmodells nicht in Frage zu stellen.

De Vos et al. (2022) definieren Reisezufriedenheit als ein Konstrukt, das sowohl die kognitive Bewertung der Fahrt als auch die während der Reise erlebten Emotionen umfasst. Ein solches differenziertes Verständnis liefert eine abschließende Bewertung durch die STS nicht, selbst wenn der Affekt zum Zeitpunkt der Bewertung sinnvoll erfasst werden könnte. Zudem können die Korrelationen der affektiven und kognitiven Bewertungen nicht als gerichteter Effekt interpretiert werden (Devos et al., 2019). Es ist zwar plausibel, dass die Dynamik der affektiven Zustände einer Reise die kognitive Bewertung dieser beeinflusst, aber Hypothesen über solche dynamischen Prozesse des Reiseerlebens sind nicht durch einmalige oder gar retrospektive Erhebungen falsifizierbar.

Experience Sampling Methoden (ESM) (z.B. Hektner, Schmidt & Csikszentmihalyi, 2007) ermöglichen eine zeitlich hochaufgelöste Erfassung der Bewertung des Reiseerlebens. Neben Abhängigkeiten der Daten wie sie in dieser Studie vorliegen, müssen weitere methodische Herausforderungen wie Autokorrelationen oder nicht zufällig fehlende Daten überwunden werden. Dynamische Strukturgleichungsmodelle (DSEM) (Asparouhov, Hamaker & Muthén, 2018; Asparouhov & Muthén, 2020) werden der Reichhaltigkeit solcher Daten durch gemeinsame Anwendung von Techniken der Faktorenanalyse, Mehrebenenmodellierung und Zeitreihenanalysen gerecht. Statt einer multiplen Regression ähneln die Modellgleichungen eher Differentialgleichungen, um die Dynamik einer oder mehrerer Variablen über die Zeit abzubilden. Unter der Annahme eines korrekt spezifizierten Messmodells, lässt sich mit dieser Methode die Reliabilität von Ein-Item-Messungen, wie sie häufig in ESM-Studien verwendet werden, prinzipiell testen. Ein Beispiel dafür ist ein Konstrukt, das sich über die

Zeit stabil verhält und nur durch spezifische Einflüsse schwankt. Abweichungen von diesem Muster könnten dann als Messfehler interpretiert und eine Reliabilität berechnet werden (Dejonckheere et al., 2022).

Darüber hinaus können Physiologische Messungen, wie die Erfassung der Hautleitfähigkeit oder der Herzrate durch EKG, die subjektiven Berichte in ES-Daten ergänzen. Eine multivariate Analyse der multimethodischen Daten kann möglicherweise Licht auf das Konstrukt Affekt und dessen Dynamik werfen. Denn ein Goldstandard zur Messung des Affekts bleibt noch aus (Meiselmann, 2021), ebenso wie eine allgemein anerkannte Definition.

DSEM ist zwar eine sich noch entwickelnde Technik, doch möglicherweise die Notwendige, um die Forschung zum Erleben, inneren Dynamiken und sogar Heuristiken in der Bewertung wiederzubeleben und voranzutreiben. Kahneman (1999) betonte selbst, dass die Forschung zum Erleben durch Methoden wie ES wieder aufgenommen werden sollte. Heutzutage ist die Durchführung von ESM-Studien durch eine weite Verbreitung der Smartphones kostengünstiger als noch vor 20 Jahren. Allerdings erfordern DSEM eine sehr hohe methodische Kompetenz. Daher bleibt es wichtig, leicht zugängliche Instrumente wie die STS kontinuierlich weiterzuentwickeln, um zuverlässige Studien in verschiedenen Kontexten und Kulturen zu ermöglichen.

Zuverlässigere Messinstrumente oder tiefergehende Erhebungsmethoden und Analysen könnten Erkenntnisse darüber liefern, ob Unterschiede in der Zufriedenheit mit verschiedenen Verkehrsmitteln die Verkehrsmittelwahl im Sinne des Mode-Choice-Cycles beeinflussen. Diese Studie zeigt jedoch, dass die überwiegende Präferenz für den Zug auf der Strecke zwischen Braunschweig und Wolfenbüttel nicht durch Unterschiede in der Reisezufriedenheit erklärt werden kann. Die genannten Gründe der Präferenzen des Zuges und des Busses sind stattdessen geprägt durch den individuellen Aufwand, der mit einem Verkehrsmittel verbunden ist. Für Personen, die näher an der Nähe der Haltestelle Rathaus

wohnen, stellt der Hauptbahnhof einen Umweg mit mehr Umstiegen und Wartezeiten dar, sodass der Bus mit direkter Route nach Wolfenbüttel mit einem geringeren Aufwand verbunden wäre.

Lediglich ein Argument deutete auf eine persönliche Präferenz für ein Verkehrsmittel hin („fahre sowieso lieber Zug“). Insgesamt können die qualitativen Begründungen so interpretiert werden, dass Reisende tendenziell das Verkehrsmittel wählen, das mit den geringsten Verhaltenskosten verbunden ist.

Die Aussagekraft der qualitativen Daten ist jedoch aufgrund der geringen Stichprobengröße begrenzt. Eine eingeschränkte statistische Power und Generalisierbarkeit aufgrund der Stichprobengröße gilt für alle Analysen dieser Studie auf Probandenebene, einschließlich der MCFA. Um Effekte mit 42 Probanden zu entdecken, müssen die Effekte relativ stark sein (Cohen, 1992; McCallum, Browne & Sugawara, 1996) oder die untersuchten Variablen präzise erfasst werden. Dagegen bieten die 387 Fahrten eine solide Grundlage, um beispielsweise den Zusammenhang zwischen berichteten Ereignissen und der Reisezufriedenheit zu untersuchen.

Doch auch diese Untersuchung kann nur eine grobe Abstraktion der Realität abbilden. Die Auswahloptionen fassen verschiedene Situationen grob zusammen. Beispielsweise werden „Störende Mitmenschen in der Umgebung“ und „Überfülltes Verkehrsmittel“ in einer Kategorie berichtet. Die tatsächliche Varianz hinter diesen Berichten kann jedoch enorm sein. Es ist unklar, ob jemand dieses Ereignis berichtet, weil sich ein fremder Fahrgast neben sie setzt, jeder zweite Sitzplatz belegt ist und die Person es deshalb bevorzugt zu stehen, das Verkehrsmittel so überfüllt ist, dass keine weiteren Fahrgäste mehr einsteigen können oder ein Fahrgast ein belegtes Brot mit Ei und Thunfisch isst. Solche Feinheiten sind durch die grobe Befragung nicht zu unterscheiden.

Ein weiteres Problem der Erfassung der Ereignisse liegt in der Struktur des ES-Fragebogens, weil die Bejahung der Frage „Ist etwas passiert?“ die Auswahl eines Ereignisses oder sogar eine offene Eingabe mit sich bringt. Bei einer Verneinung hingegen fällt dieser zusätzliche Aufwand weg. Da die Probanden häufig den gleichen Fragebogen ausfüllen, lernen sie schnell, dass eine Verneinung weniger Aufwand

bedeutet, was die Validität der Berichte potentiell einschränkt (Meade & Craig, 2012; Krosnick, 1991). Aus diesem Grund wurde nicht die Häufigkeit der Ereignisse berücksichtigt, sondern nur ob sie mindestens einmal in einer Fahrt berichtet wurden.

Die Interpretationskraft ist aufgrund der explorativen Natur der Analysen ohnehin eingeschränkt. Die signifikanten Ergebnisse der explorativen Analysen überlebten vermutlich keine Alpha-Fehler Korrektur des multiplen Hypothesentestens (z.B. Miller, 1981) und sind deshalb unter Vorsicht zu interpretieren.

Trotz dieser methodischen Einschränkungen zeigt diese Arbeit, dass mit nur 42 Probanden durch Messwiederholungen eine faktorenanalytische Modellierung möglich ist, insbesondere zur Messung situativer Konstrukte ist es entscheidend, dass diese über retrospektive Befragungen oder hypothetische Szenarien hinaus im realen Kontext erhoben werden. Die MCFA erweist sich hierfür als wertvolle Methode und ermöglicht die Analyse der fluktuierenden und stabilen Anteile eines Konstrukts. Zudem ermöglicht die Implementierung in der frei zugänglichen Software *R lavaan* (2024) die kostenfreie Durchführung dieser Analyse. Ein wesentlicher Beitrag dieser Arbeit liegt in der Interpretation der Faktorenstrukturen auf verschiedenen Ebenen. Sowohl Faktorenanalyse als auch Mehrebenenmodellierung sind methodisch ausgereifte Konzepte, doch die Interpretation der Ergebnisse ihrer gemeinsamen Anwendung bedarf weiterer Forschung und Simulationen (Stapleton et al., 2016; Jak, 2019).

Damit Verkehrsbetriebe von Bewertungen des Reiseerlebens effektiv profitieren können, müssen Skalen zur Erfassung der Reisezufriedenheit weiterentwickelt werden. Selbst bei einem optimierten Messinstrument bleibt es jedoch herausfordernd, den Einfluss der bewertenden Person von der tatsächlichen Reise zu differenzieren. Eine verbesserte Skala sollte daher idealerweise von einer großen Gruppe Probanden ausgefüllt werden, die gleichzeitig dieselbe Reise erleben. Dies ermöglicht eine objektivere Beurteilung externer Faktoren, wie eines überfüllten Verkehrsmittels. Die Erfassung mehrerer solcher Reisen ergäbe eine hierarchische Datenstruktur, bei der die Bewertungen der Probanden

Abhängigkeiten aufgrund der gemeinsam erlebten Reise aufweisen. Eine MCFA, die Probanden auf der Within-Ebene und verschiedene Fahrten auf der Between-Ebene modelliert, könnte einen Faktor auf der Between-Ebene offenbaren, der die latente Servicequalität eines Verkehrsmittels oder einer spezifischen Reise darstellt, unabhängig individueller Wahrnehmungen und Bewertungen. Diese latente Servicequalität könnte als Grundlage für gezielte Verbesserungsmaßnahmen des Reiseerlebens im ÖPNV dienen.

Die Messung latenter Konstrukte bleibt eine fundamentale Herausforderung in der psychologischen Forschung. Vielversprechende Theorien scheitern oft nicht an ihrer inhaltlichen Tiefe, sondern an einer unzureichenden Operationalisierung der gemessenen Konstrukte gepaart mit naiven Erhebungsinstrumenten, die keine überzeugende Empirie bieten können. Es ist vergeblich, die wahren Zusammenhänge der Variablen dieser Welt zu erforschen oder gar in Proportionsverhältnissen darzustellen, solange sie nicht valide und reliabel erfasst und in sinnvollen Zahlen ausgedrückt werden können.

Literaturverzeichnis

- Abenoza, R. F., Cats, O., & Susilo, Y. O. (2019). How does travel satisfaction sum up? An exploratory analysis in decomposing the door-to-door experience for multimodal trips. *Transportation*, 46(5), 1615-1642.
- Aitken, A. C. (1936). On Least Squares and Linear Combination of Observations. *Proceedings of the Royal Society of Edinburgh*, 55, 42–48.
- Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2), 179–211.
- Ariely, D., & Loewenstein, G. (2000). When does duration matter in judgment and decision making?. *Journal of experimental psychology: general*, 129(4), 508.
- Asparouhov, T., Hamaker, E.L. & Muthén, B. (2018). Dynamic structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, 25(3), 359-388.
- Asparouhov, T. & Muthén, B. (2020). Comparison of models for the analysis of intensive longitudinal data. *Structural Equation Modeling: A Multidisciplinary Journal*, 27(2) 275-297.
- Barrett, L. F. (2006). Are emotions natural kinds?. *Perspectives on psychological science*, 1(1), 28-58.
- Barrett, L. F. (2017). The theory of constructed emotion: an active inference account of interoception and categorization. *Social cognitive and affective neuroscience*, 12(1), 1-23.
- Bollen, K. A. (1980). *Issues in the comparative measurement of political democracy*. *American Sociological Review*, 45(3), 370–390.
- Borg, I., & Staufienbiel, T. (2007). Theorien und Methoden der Skalierung. *Aufl., Bern*.
- Bosch, E., Luther, R., & Ihme, K. (2024). Travel Experience in Public Transport Dataset. *arXiv preprint arXiv:2410.02792*.

Bundesministerium für Digitales und Verkehr (2023). *Förderprogramm zur Stärkung des ÖPNV*. Zugriff am 21.10.2024 auf <https://www.bmvi.de/DE/Themen/Mobilitaet/OEPNV/foerderprogramm-staerkung-oepnv>

Blyth, C. (1972). On Simpson's paradox and the sure-thing principle. *Journal of the American Statistical Association*, 67, 364–366.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological bulletin*, 56(2), 81.

Cohen, J. (1992). Statistical Power Analysis. *Current Directions in Psychological Science*, 1(3), 98-101.

Costa, P. T., & McCrae, R. R. (1980). Influence of extraversion and neuroticism on subjective well-being: happy and unhappy people. *Journal of personality and social psychology*, 38(4), 668.

Dejonckheere, E., Demeyer, F., Geusens, B., Piot, M., Tuerlinckx, F., Verdonck, S., & Mestdagh, M. (2022). Assessing the reliability of single-item momentary affective measurements in experience sampling. *Psychological assessment*, 34(12), 1138.

De Vos, J., Schwanen, T., Van Acker, V., & Witlox, F. (2015). How satisfying is the Scale for Travel Satisfaction?. *Transportation Research Part F: Traffic Psychology and Behaviour*, 29, 121-130.

De Vos, J., Schwanen, T., Van Acker, V., & Witlox, F. (2019). Do satisfying walking and cycling trips result in more future trips with active travel modes? An exploratory study. *International Journal of Sustainable Transportation*, 13(3), 180-196.

De Vos, J., Singleton, P. A., & Gärling, T. (2022). From attitude to satisfaction: introducing the travel mode choice cycle. *Transport Reviews*, 42(2), 204-221.

Diener, E. D., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The satisfaction with life scale. *Journal of personality assessment*, 49(1), 71-75.

Domjan, M. (2013). *The principles of learning and behavior* (7th ed.). Wadsworth/Cengage Learning.

Duncan, T. E., Duncan, S. C., Alpert, A., Hops, H., Stoolmiller, M., & Muthen, B. (1997). Latent variable modeling of longitudinal and multilevel substance use data. *Multivariate Behavioral Research, 32*(3), 275-318.

Dunn, E. C., Masyn, K. E., Jones, S. M., Subramanian, S. V., Koenen, K. C., & Earls, F. (2015). Measuring psychosocial environments using individual responses: An application of multilevel factor analysis to examining students in schools. *Prevention Science, 16*(5), 718–733.

Ettema, D., Gärling, T., Eriksson, L., Friman, M., Olsson, L. E., & Fujii, S. (2011). Satisfaction with travel and subjective well-being: Development and test of a measurement tool. *Transportation Research Part F: Traffic Psychology and Behaviour, 14*(3), 167-175.

Fisher, R. A. (1925). *Statistical methods for research workers*. Edinburgh: Oliver and Boyd.

Friman, M., Fujii, S., Ettema, D., Gärling, T., & Olsson, L. E. (2013). Psychometric analysis of the satisfaction with travel scale. *Transportation Research Part A: Policy and Practice, 48*, 132-145.

Geldhof, G. J., Preacher, K. J., & Zyphur, M. J. (2014). Reliability estimation in a multilevel confirmatory factor analysis framework. *Psychological methods, 19*(1), 72.

Glasgow, T. E., Geller, E. S., Le, H. T., & Hankey, S. (2018). Travel mood scale: Development and validation of a survey to measure mood during transportation. *Transportation research part F: traffic psychology and behaviour, 59*, 318-329.

Gulliksen, H. (1950). *Theory of mental tests*. John Wiley & Sons.

Hektner, J. M., Schmidt, J. A., & Csikszentmihalyi, M. (2007). *Experience sampling method*. SAGE Publications.

Henn, L., Otto, S., & Kaiser, F. G. (2020). Positive spillover: The result of attitude change. *Journal of Environmental Psychology, 69*, 101429.

Hu, L.-t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1–55.

Hoffman, L. (2015). *Longitudinal analysis: Modeling within-person fluctuation and change*. Routledge.

Hoffman, L., & Walters, R. W. (2022). Catching up on multilevel modeling. *Annual Review of Psychology*, 73(1), 659-689.

Hoyle, R. H. (Ed.). (2012). *Handbook of structural equation modeling*. Guilford press.

Hox, J. J., & Maas, C. J. M. (2002). *Sample sizes for multilevel modeling*. In J. Blasius, J. J. Hox, E. de Leeuw, & P. Schmidt (Hrsg.). *Social science methodology in the new millennium: Proceedings of the fifth international conference on logic and methodology* (2). Leske + Budrich.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal*, 6(1), 1-55.

Huang, F. (2017). *Conducting multilevel confirmatory factor analysis using R*. Zugriff am 21.10.2024 auf <https://francish.netlify.app/docs/MCFainRHUANG.pdf> und <https://github.com/flh3/mcfa>.

Jak, S. (2019). Cross-level invariance in multilevel factor models. *Structural Equation Modeling: A Multidisciplinary Journal*, 26(4), 607-622.

Janke, S., & Glöckner-Rist, A. (2014). Deutsche version der satisfaction with life scale (SWLS). *Zusammenstellung sozialwissenschaftlicher Items und Skalen*. Mannheim: GESIS.

Julian, M. W. (2001). The consequences of ignoring multilevel data structures in nonhierarchical covariance modeling. *Structural equation modeling*, 8(3), 325-352.

Jöreskog, K.G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, 34, 183-202.

Jöreskog, K.G. (1971). Statistical analysis of sets of congeneric tests. *Psychometrika*, 36, 109-133.

Kahneman, D. (1999). Objective happiness. In D. Kahneman, E. Diener, & N. Schwarz (Hrsg.), *Well-being: The foundations of hedonic psychology* (pp. 3–25). Russell Sage Foundation.

Kahneman, D. (2000). Evaluation by moments: Past and future. *Choices, values, and frames*, 693-708. Cambridge University Press.

Kahneman, D., & Frederick, S. (2005). A model of heuristic judgment. *The Cambridge handbook of thinking and reasoning*, 267, 293.

Kaiser, F. G. (1998). A general measure of ecological behavior. *Journal of applied social psychology*, 28(5), 395-422.

Kaiser, F. G., Byrka, K., & Hartig, T. (2010). Reviving Campbell's paradigm for attitude research. *Personality and Social Psychology Review*, 14(4), 351-367.

Kaiser, F. G. (2020). GEB-50: General ecological behavior scale. *PsychArchives*.

Kaiser, F. G., Henn, L., & Marschke, B. (2020). Financial rewards for long-term environmental protection. *Journal of Environmental Psychology*, 68, 101411.

Kaiser, F. G. (2021). Climate change mitigation within the Campbell paradigm: doing the right thing for a reason and against all odds. *Current Opinion in Behavioral Sciences*, 42, 70-75.

Kaiser, F. G., Kibbe, A., & Hentschke, L. (2021). Offsetting behavioral costs with personal attitudes: A slightly more complex view of the attitude-behavior relation. *Personality and Individual Differences*, 183, 111158.

Koch, G. G. (1983). Intraclass correlation coefficient. *Encyclopedia of Statistical Sciences*, 4, 212-217.

Kolenikov, S., & Bollen, K. A. (2012). Testing negative error variances: Is a Heywood case a symptom of misspecification?. *Sociological Methods & Research*, 41(1), 124-167.

Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied cognitive psychology*, 5(3), 213-236.

Kyriazos, T. A. (2019). Applied psychometrics: the modeling possibilities of multilevel confirmatory factor analysis (MLV CFA). *Psychology*, 10(06), 777.

Lee, H., Calvin, K., Dasgupta, D., Krinner, G., Mukherji, A., Thorne, P., & Park, Y. (2023). *Climate change 2023: Synthesis report, summary for policymakers. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Core Writing Team, H. Lee & J. Romero (Hrsg.). Intergovernmental Panel on Climate Change.

Lim, T., Thompson, J., Tian, L., & Beck, B. (2023). A transactional model of stress and coping applied to cyclist subjective experiences. *Transportation research part F: traffic psychology and behaviour*, 96, 155-170.

Lim, T., Thompson, J., Pearson, L., Odgers, J. C., & Beck, B. (2024). Effects of within-trip subjective experiences on travel satisfaction and travel mode choice: A conceptual framework. *Transportation Research Part F: Traffic Psychology and Behaviour*, 104, 201-216.

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, 1(2), 130-149.

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological methods*, 17(3), 437.

Meiselman, H. L. (2021). Emotion measurement: integrative summary. In *Emotion measurement* (2. Ed.) (pp. 645-697). Woodhead Publishing.

Miller, R. G. (1981). *Simultaneous statistical inference* (2nd ed.). Springer.

Mischel, W. (1969). Continuity and change in personality. *American Psychologist*, 24(11), 1012-1018.

Monteiro, J., Sousa, N., Coutinho-Rodrigues, J., & Natividade-Jesus, E. (2024). Challenges ahead for sustainable cities: An urban form and transport system review. *Energies*, *17*(2), 409.

Movisens GmbH. (2023). Movisensxs (Version 1.5.23). <https://www.movisens.com/en/products/movisensxs/>.

Moosbrugger, H., and Kelava, A. (2020). *Testtheorie und Fragebogenkonstruktion*. Springer.

Muthén, B. O. (1989). Latent variable modeling in heterogeneous populations. *Psychometrika*, *54*, 557-585.

Muthén, B. O. (1994). Multilevel Covariance Structure Analysis. *Sociological Methods & Research*, *22*(3), 376-398.

Muthén, B., & Asparouhov, T. (2011). Beyond multilevel regression modeling: Multilevel analysis in a general latent variable framework. In *Handbook of advanced multilevel analysis* (pp. 15-40). Routledge.

Muthén, L. K., & Muthén, B. O. (2024). *Mplus user's guide* (8.8 ed.).

Muthén, B. O., & Satorra, A. (1995). Complex sample data in structural equation modeling. *Sociological methodology*, 267-316.

Pearl, J. (2022). Comment: understanding Simpson's paradox. In H. Geffner, R. Dechter, J. Y. Halpern (Eds.), *Probabilistic and Causal inference: The works of Judea Pearl* (pp. 399-412). Association for Computing Machinery

Olsson, L. E., Friman, M., Pareigis, J., & Edvardsson, B. (2012). Measuring service experience: Applying the satisfaction with travel scale in public transport. *Journal of Retailing and Consumer Services*, *19*(4), 413-418.

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). *Common method biases in behavioral research: A critical review of the literature and recommended remedies*. *Journal of Applied Psychology*, *88*(5), 879-903.

Rammstedt, B., Kemper, C., Klein, M. C., Beierlein, C., & Kovaleva, A. (2013). Eine kurze Skala zur Messung der fünf Dimensionen der Persönlichkeit: big-five-inventory-10 (BFI-10). *Methoden, Daten, Analysen (mda)*, 7(2).

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Danish Institute for Educational Research: Copenhagen.

Raykov, T., & Marcoulides, G. A. (2016). On the relationship between classical test theory and item response theory: From one to the other and back. *Educational and psychological measurement*, 76(2), 325-338.

R Core Team. (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Redelmeier, D. A., & Kahneman, D. (1996). Patients' memories of painful medical treatments: Real-time and retrospective evaluations of two minimally invasive procedures. *pain*, 66(1), 3-8.

Redelmeier, D. A., Katz, J., & Kahneman, D. (2003). Memories of colonoscopy: a randomized trial. *Pain*, 104(1-2), 187-194.

Reise, S. P., Bonifay, W. E., & Haviland, M. G. (2013). Scoring and modeling psychological measures in the presence of multidimensionality. *Journal of personality assessment*, 95(2), 129-140.

Rosenberg, M. J., & Hovland, C. I. (1960). Cognitive, affective, and behavioral components of attitudes. In C. I. Hovland & M. J. Rosenberg (Eds.), *Attitude organization and change: An analysis of consistency among attitude components* (pp. 1–14). Yale University Press.

Rosseel, Y. (2024). *lavaan: Latent variable analysis*. R package version 0.6-13.

Russell, J. A. (1980). A circumplex model of affect. *Journal of personality and social psychology*, 39(6), 1161.

Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye & C. C. Clogg (Hrsg.), *Latent variables analysis: Applications for developmental research* (pp. 399–419). Sage Publications.

Scherer, K. R. (2001). Appraisal considered as a process of multilevel sequential checking. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, methods, research* (pp. 92–120). Oxford University Press.

Scherer, K. R. (2019). Towards a prediction and data driven computational process model of emotion. *IEEE Transactions on Affective Computing, 12*(2), 279-292.

Shapira, A.A., Pansky, A. Cognitive and metacognitive determinants of eyewitness memory accuracy over time. *Metacognition Learning 14*, 437–461.

Singleton, P. A. (2019). Validating the Satisfaction with Travel Scale as a measure of hedonic subjective well-being for commuting in a US city. *Transportation research part F: traffic psychology and behaviour, 60*, 399-414.

Skinner, B. F. (1938). *The behavior of organisms: an experimental analysis*. Appleton-Century.

Smith, O. (2017). Commute well-being differences by mode: Evidence from Portland, Oregon, USA. *Journal of Transport & Health, 4*, 246-254.

Sober, E. (1981). The principle of parsimony. *British Journal for the Philosophy of Science, 32*(2), 145-156.

Stapleton, L. M. (2006). Using multilevel structural equation modeling techniques with complex sample data. *Structural equation modeling: A second course*, 345-383.

Stapleton, L. M., Yang, J. S., & Hancock, G. R. (2016). Construct meaning in multilevel settings. *Journal of Educational and Behavioral Statistics, 41*, 481-520.

Steinhorst, J., & Klöckner, C. A. (2018). Effects of monetary versus environmental information framing: Implications for long-term pro-environmental behavior and intrinsic motivation. *Environment and Behavior, 50*(9), 997-1031.

Steyer, R., Eid, M. (2001). Grundbegriffe der Klassischen Testtheorie. In: Messen und Testen. Springer.

Snijders, T., & Bosker, R. (2012). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. Sage Publishers.

Thorndike, E. L. (1933). A proof of the law of effect. *Science, 77*(1989), 173-175.

Västfjäll, D., Friman, M., Gärling, T., & Kleiner, M. (2002). The measurement of core affect: A Swedish self-report measure derived from the affect circumplex. *Scandinavian Journal of Psychology, 43*(1), 19-31.

Weidman, A. C., Steckler, C. M., & Tracy, J. L. (2017). The jingle and jangle of emotion assessment: Imprecise measurement, casual scale usage, and conceptual fuzziness in emotion research. *Emotion, 17*(2), 267.

Wetzel, E., & Frick, S. (2020). Comparing the validity of trait estimates from the multidimensional forced-choice format and the rating scale format. *Psychological Assessment, 32*(3), 239.

Wilson, M. (2023). *Constructing measures: An item response modeling approach*. Routledge.

Winer, B. J., Brown, D.R. & Michels, K.M. (1991). *Statistical Principles in Experimental Design*. McGraw-Hill.

Yu, C. H. (2020). *Objective measurement: How Rasch modeling can simplify and enhance your assessment*. In M. Khine (Ed.), *Rasch Measurement* (pp. 67-82). Springer.

Yuan, K.-H., & Bentler, P. M. (2000). Three likelihood-based methods for mean and covariance structure analysis with nonnormal missing data. *Sociological Methodology, 30*(1), 165-200.

Anhang A

Die unstandardisierten Kovarianzmatrizen SPW und c SB wurden zur individuellen Schätzung der Within- und Between- Messmodelle verwendet und können genutzt werden, um selbst Messmodelle zu testen. N = 387 G = 42

# Item - Faktor	1	2	3	4	5	6	7	8	9
Pooled-Within									
Kovarianzmatrix									
1 - PD	1.17								
2 - PD	.58	1.09							
3 - PD	.70	.54	1.61						
4 - PA	.17	.27	.33	1.52					
5 - PA	.27	.32	.33	.57	1.09				
6 - PA	.45	.34	.57	.37	.41	1.09			
7 - CE	.44	.38	.43	.20	.36	.45	.86		
8 - CE	.36	.29	.33	.18	.41	.49	.60	.93	
9 - CE	.47	.34	.41	.21	.40	.54	.61	.72	1.56
Angepasste Between									
Kovarianzmatrix									
1 - PD	.48								
2 - PD	.51	.60							
3 - PD	.52	.59	.75						
4 - PA	.47	.49	.48	.86					
5 - PA	.37	.40	.33	.47	.56				
6 - PA	.43	.48	.51	.54	.54	.61			
7 - CE	.33	.39	.35	.27	.27	.27	.53		
8 - CE	.37	.44	.39	.39	.39	.30	.53	.59	
9 - CE	.35	.46	.45	.21	.21	.24	.34	.32	.50

Anhang B

Das ungewöhnliche Within-Modell birgt die Herausforderung, dass Item 6 direkt auf dem hierarchischen Faktor lädt und die übrigen Items über PA bzw. CE auf dem hierarchischen Faktor laden. R-Pakete neben *lavaan* wie *psych* oder *semTools* bieten für ein solches Modell keine vorbereitete Funktion zur Reliabilitätsberechnung. Deshalb wird im Folgenden die eigenständige Berechnung der Reliabilitätskoeffizienten dargestellt. Schermelleh-Engel und Gäde in Moosbrugger & Kelava (2020, Kap. 15 S. 337-364) liefern detaillierte Erklärungen zur folgenden modellbasierten Reliabilitätsschätzung.

Die Modellgleichungen des MMM lauten:

$$y_1 = v_1 + \lambda_{w1} \cdot PD_w + \epsilon_{w1} + \lambda_{b1} \cdot \eta_b + \epsilon_{b1}$$

$$y_2 = v_2 + \lambda_{w2} \cdot PD_w + \epsilon_{w2} + \lambda_{b2} \cdot \eta_b + \epsilon_{b2}$$

$$y_3 = v_3 + \lambda_{w3} \cdot PD_w + \epsilon_{w3} + \lambda_{b3} \cdot \eta_b + \epsilon_{b3}$$

$$PD_w = \beta_1 \cdot TS_w + \zeta_{w1}$$

$$y_6 = v_6 + \lambda_{w6} \cdot TS_w + \epsilon_{w6} + \lambda_{b6} \cdot \eta_b + \epsilon_{b6}$$

$$y_7 = v_7 + \lambda_{w7} \cdot CE_w + \epsilon_{w7} + \lambda_{b7} \cdot \eta_b + \epsilon_{b7} + \text{Cov}(\epsilon_{b7}, \epsilon_{b8})$$

$$y_8 = v_8 + \lambda_{w8} \cdot CE_w + \epsilon_{w8} + \lambda_{b8} \cdot \eta_b + \epsilon_{b8} + \text{Cov}(\epsilon_{b7}, \epsilon_{b8})$$

$$y_9 = v_9 + \lambda_{w9} \cdot CE_w + \epsilon_{w9} + \lambda_{b9} \cdot \eta_b + \epsilon_{b9}$$

$$CE_w = \beta_2 \cdot TS_w + \zeta_{w2}$$

Legende:

y_1, y_2, \dots, y_9 : Antworten der Items der STS

v_1, v_2, \dots, v_9 : Intercepts der jeweiligen Items

$\lambda_{w1}, \lambda_{w2}, \dots, \lambda_{w9}$: Faktorladungen der Items auf ihre jeweiligen Faktoren auf der Within-Ebene

$\lambda_{b1}, \lambda_{b2}, \dots, \lambda_{b9}$: Faktorladungen der Items auf ihren Faktor auf der Between-Ebene

$\epsilon_{w1}, \epsilon_{w2}, \dots, \epsilon_{w9}$: Residuen auf der Within-Ebene

$\epsilon_{b1}, \epsilon_{b2}, \dots, \epsilon_{b9}$: Residuen auf der Between-Ebene

ζ_{w1}, ζ_{w2} : Disturbances (Störgrößen) der Within-Faktoren 1. Ordnung

PD_w: Faktor „Positive Deaktivierung“ auf der Within-Ebene (1. Ordnung)

CE_w: Faktor „Kognitive Evaluation“ auf der Within-Ebene (1. Ordnung)

TS_w: Faktor „Reisezufriedenheit“ auf der Within-Ebene (2. Ordnung)

β_1, β_2 : Strukturkoeffizienten der Faktoren PD_w und CE_w auf den Faktor TS_w

η_b : Between-Faktor, Neigung zu positiven oder negativen Bewertungen

$\text{Cov}(\epsilon_{b7}, \epsilon_{b8})$: Kovarianz der Residuen zwischen den Items 7 und 8 auf der Between-Ebene

Im Modell wurden ϵ_{b1} und ϵ_{b2} , wie in Abbildung 6 beschrieben, auf .005 fixiert. λ_{w1} , λ_{w7} und β_1 sind auf 1 fixiert zur Skalierung der latenten Variablen und y_7 und y_8 enthalten die zugelassenen kovariierenden Residuen.

Daraus folgt für die modellimplizierte Varianz der Summenwerte der Indikatorvariablen (für Details Kapitel 15 und Kapitel 24 und in Moosbrugger & Kelava, 2020):

$$\begin{aligned} \text{Var}(Y) = & \left(\sum_{i=1}^3 \lambda_{w,i} \cdot \beta_1 + \lambda_{w,6} + \sum_{i=7}^9 \lambda_{w,i} \cdot \beta_2 \right)^2 \cdot \text{Var}(TS_w) + \left(\sum_{i=1}^3 \lambda_{w,i} \right)^2 \cdot \text{Var}(PD_w) + \left(\sum_{i=7}^9 \lambda_{w,i} \right)^2 \\ & \cdot \text{Var}(CE_w) + \sum_{i \in \{1,2,3,6,7,8,9\}} \epsilon_{w,i} + \left(\sum_{i \in \{1,2,3,6,7,8,9\}} \lambda_{b,i} \right)^2 \cdot \text{Var}(\eta_b) + \sum_{i \in \{1,2,3,6,7,8,9\}} \epsilon_{b,i} \\ & + 2 \cdot \text{Cov}(\epsilon_{b,7}, \epsilon_{b,8}) \end{aligned}$$

Reliabilität meint das Verhältnis von „wahrer“ Varianz zur Gesamtvarianz (wahre + Messfehler). Geldhof (2014) beschreibt Composite Reliability getrennt für die Within- und Between Anteile der gesamten Varianz.

Die Between-Seite des Messmodells ist nur im Wesentlichen eindimensional aufgrund der Kovarianz zwischen den Residuen der Items 7 und 8. Da diese Kovarianz inhaltlich relevant ist, sollte sie in die Reliabilitätsberechnung miteinbezogen werden. Das geeignete Reliabilitätsmaß für dieses eindimensionale, tau-kongenerische Modell mit einer zugelassenen Residualkovarianz ist Bollen's Omega (ω^*) (Bollen, 1980; Raykov & Marcoulides, 2016):

$$\omega_b^* = \frac{(\sum_{i \in \{1,2,3,6,7,8,9\}} \lambda_{b,i})^2 \cdot \text{Var}(\eta_b)}{(\sum_{i \in \{1,2,3,6,7,8,9\}} \lambda_{b,i})^2 \cdot \text{Var}(\eta_b) + \sum_{i \in \{1,2,3,6,7,8,9\}} \epsilon_{b,i} + 2 \cdot \text{Cov}(\epsilon_{b,7}, \epsilon_{b,8})}$$

Durch Einsetzen der standardisierten Modellparameter aus Abbildung 4 erhalten wir:

$$\omega_b^* = \frac{(0.994 + 0.996 + 0.893 + 0.774 + 0.698 + 0.707 + 0.779)^2 \cdot 1}{34.117 \cdot 1 + (0.011 + 0.009 + 0.203 + 0.401 + 0.513 + 0.500 + 0.392) + 2 \cdot 0.934}$$

$$\omega_b^* = \frac{34.117}{34.117 + 2.029 + 1.868} = 0.897$$

Aufgrund der hierarchischen Struktur des Within-Modells sind wenigstens zwei Reliabilitätsmaße interessant.

OmegaTotal (ω_{TW}) beschreibt den Anteil der modellierten Varianz der Indikatoren, der auf die Faktoren zurückzuführen ist, an der gesamten Varianz.

OmegaTotal:

$$\omega_{TW} = \frac{(\sum_{i=1}^3 \lambda_{w,i} \cdot \beta_1 + \lambda_{w,6} + \sum_{i=7}^9 \lambda_{w,i} \cdot \beta_2)^2 \cdot \text{Var}(TS) + (\sum_{i=1}^3 \lambda_{w,i})^2 \cdot \text{Var}(PD) + (\sum_{i=7}^9 \lambda_{w,i})^2 \cdot \text{Var}(CE)}{(\sum_{i=1}^3 \lambda_{w,i} \cdot \beta_1 + \lambda_{w,6} + \sum_{i=7}^9 \lambda_{w,i} \cdot \beta_2)^2 \cdot \text{Var}(TS) + (\sum_{i=1}^3 \lambda_{w,i})^2 \cdot \text{Var}(PD) + (\sum_{i=7}^9 \lambda_{w,i})^2 \cdot \text{Var}(CE) + \sum_{i \in \{1,2,3,6,7,8,9\}} \epsilon_{w,i}}$$

$$\omega_{TW} = \frac{\left(\begin{array}{l} 0.774 \cdot 0.769 + 0.641 \cdot 0.769 + 0.663 \cdot 0.769 \\ + 0.709 + 0.837 \cdot 0.798 + 0.800 \cdot 0.798 + 0.730 \cdot 0.798 \end{array} \right)^2 \cdot 1 + (0.774 + 0.641 + 0.663)^2 \cdot 0.409 + (0.837 + 0.800 + 0.730)^2 \cdot 0.364}{17.605 + (0.774 + 0.641 + 0.663)^2 \cdot 0.409 + (0.837 + 0.800 + 0.730)^2 \cdot 0.364 + (0.400 + 0.589 + 0.560 + 0.498 + 0.300 + 0.360 + 0.467)}$$

$$\omega_{TW} = \frac{17.605 + 1.766 + 2.043}{17.605 + 1.766 + 2.043 + 3.174} = 0.871$$

Omega hierarchisch (ω_{HW}) beschreibt den Anteil des hierarchischen Faktors 2. Ordnung an der gesamten modellierten Varianz. Nur die durch den Faktor TS_w erklärte Varianz der Items wird dementsprechend als wahre Varianz betrachtet.

$$\omega_{HW} = \frac{(\sum_{i=1}^3 \lambda_{w,i} \cdot \beta_1 + \lambda_{w,6} + \sum_{i=7}^9 \lambda_{w,i} \cdot \beta_2)^2 \cdot \text{Var}(TS_w)}{(\sum_{i=1}^3 \lambda_{w,i} \cdot \beta_1 + \lambda_{w,6} + \sum_{i=7}^9 \lambda_{w,i} \cdot \beta_2)^2 \cdot \text{Var}(TS_w) + (\sum_{i=1}^3 \lambda_{w,i}^2) \cdot \text{Var}(PD_w) + (\sum_{i=7}^9 \lambda_{w,i}^2) \cdot \text{Var}(CE_w) + \sum_{i \in \{1,2,3,6,7,8,9\}} \epsilon_{w,i}}$$

$$\omega_{HW} = \frac{17.605}{17.605 + 1.766 + 2.043 + 3.174} = \frac{17.605}{24.588} = 0.716$$

Eigenständigkeitserklärung

Ich erkläre, dass ich die vorliegende Arbeit selbstständig und ohne Benutzung anderer als den angegebenen Hilfsmitteln angefertigt und alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten oder unveröffentlichten Schriften entnommen wurden, als solche kenntlich gemacht habe. Diese Arbeit wurde nicht, auch nicht auszugsweise, bereits für eine andere Prüfung angefertigt.

Magdeburg, 19.11.2024 S. Bohm